

1-1-2009

Improving surgical training : maximising the benefits of surgical simulation

Emily Boyle

Royal College of Surgeons in Ireland

Citation

Boyle, E. Improving surgical training : maximising the benefits of surgical simulation. [MD Thesis]. Dublin: Royal College of Surgeons in Ireland; 2009.

This Thesis is brought to you for free and open access by the Theses and Dissertations at e-publications@RCSI. It has been accepted for inclusion in MD theses by an authorized administrator of e-publications@RCSI. For more information, please contact epubs@rcsi.ie.

— Use Licence —

Creative Commons Licence:



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/).

Improving surgical training – maximising the benefits of surgical simulation

Emily Boyle, MRCSI

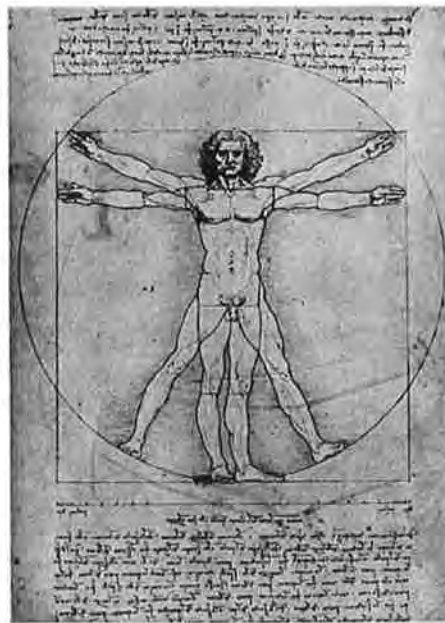
**National Surgical Training Centre,
Royal College of Surgeons in Ireland,
121 St Stephen's Green,
Dublin 2.**



Doctorate of Medicine

December 2009

**National University of Ireland
(Royal College of Surgeons in Ireland)**



Supervisor: Prof ADK Hill

**Co-supervisors –
Mr Paul C Neary
Ms Deborah McNamara**

Candidate Declaration

I declare that this thesis, which I submit to RCSI for examination in consideration of the award of a higher degree MD, is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in RCSI or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed Emily Bayle

RCSI Student Number _____

Date 31/03/2010

Table of Contents

Acknowledgements	i
Thesis Abstract	ii
Chapter 1 Background	1
1.1 Brief history of surgery	2
1.1.1 Surgery in ancient times	2
1.1.2 Surgery as a profession	3
1.1.3 Important developments in surgery	5
1.2 Endoscopy overview	7
1.21 Early endoscopy	7
1.22 Developments in endoscopy	7
1.23 Cameras in endoscopy	10
1.24 Modern era of colonoscopy	10
1.25 Complications of colonoscopy	11
1.3 Laparoscopy overview	12
1.31 History of laparoscopy	12
1.32 Laparoscopy outside of general surgery	14
1.33 Modern era of laparoscopy	14
1.34 Hand-assisted laparoscopic colectomy (HALC)	17
1.4 Endovascular surgery overview	19
1.4.1 Overview	19
1.4.2 Early developments in endovascular surgery	19
1.4.3 Current practice in endovascular surgery	21
1.5 Surgical training	23
1.5.1 Historical surgical training	23
1.5.2 Halsted and the residency system	25
1.5.3 Current Irish structure in surgical training	28
1.5.4 Current issues in surgical training	30
1.5.4.1 Work hour limitations	30
1.5.4.2 Increasing complexity in surgery	34
1.5.4.3 Economic factors	35
1.5.4.4 Ethical issues	36
1.5.4.5 Lack of standards	37

1.5.4.6 Competency	39
1.5.4.7 Feminisation of surgery	41
1.5.5 Training platforms	42
1.6 Simulation	46
1.6.1 Simulation in the aviation industry	47
1.6.2 Virtual Reality simulation	49
1.6.3 VR simulation in surgery	49
1.6.4 Developments in surgical simulators	50
1.6.5 Augmented reality simulation	51
1.6.6 Simulators as training tools	53
1.6.7 Effectiveness of simulation – skills transfer to the operating room	55
1.7 Teaching surgical skills	57
1.7.1 Theories of learning	58
1.7.2 Adult learning	60
1.7.3 Theories of motor learning – Fitts and Posner	61
1.7.4 Model of skills acquisition – Dreyfus and Dreyfus model	62
1.7.5 Designing a skills training curriculum	64
1.7.5.1 Task deconstruction	64
1.7.5.2 Practice.	65
1.7.5.3 Feedback	65
1.7.5.4 Assessment	66
1.7.5.5 Curriculum	67
1.8 Assessment of surgical performance.	68
1.8.1 Academic assessment	68
1.8.2 Human factors assessment	69
1.8.3 Technical assessment	70
1.8.3.1 Assessment of real operative performance –outcome	70
1.8.3.2 Assessment of real operative performance - directly observed assessment	72
1.8.3.3 Bench model assessment	73
1.8.3.4 Hand motion analysis	73
1.8.3.5 Simulator assessment	74

Chapter 2 Materials and methods overview	76
2.1 Overall aims of the studies	76
2.2 Subjects	78
2.3 Methodological issues	79
2.3.1 Ethical approval	79
2.3.2 Training	80
2.3.3 Assessment and endpoints	80
2.3.4 Statistical analysis	81
2.4 Description of the simulators used in the studies	81
2.4.1 GI Mentor (Simbionix, Cleveland, USA.)	81
2.4.2 ProCedicus VIST™ (Vascular Interventional Surgical Trainer) system (Mentice, Gothenberg, Sweden)	84
2.4.3 The ProMIS Laparoscopic Colectomy simulator (Haptica, Dublin, Ireland)	86
2.4.4 LapSim, (Surgical Science, Gothenberg, Sweden)	94
2.4.5 Nintendo Wii video game	95
2.5 Description of fundamental skills testing	97
2.5.1 Visuo-spatial	97
2.5.2 Psychomotor testing	99
2.5.3 Perception	99
Chapter 3 Improving surgical training: feedback reduces error scores in simulated hand-assisted laparoscopic colectomy	101
3.1 Background	102
3.2 Materials and Methods	105
3.2.1 Ethical approval	105
3.2.2 Subjects	105
3.2.3 Simulator	105
3.2.4 Study Procedures	105
3.2.4.1 Teaching session	106
3.2.4.2 Simulated HALC cases	107
3.2.4.3 Feedback – groups	107

3.2.5 Formal assessment	108
3.2.6 Sub-study – skills transfer	109
3.2.7 Statistical analysis	109
3.3 Results	109
3.3.1 Demographics	109
3.3.2 Overall results, learning curve	110
3.3.3 Results for groups 1 and 2 – effect of feedback	112
3.3.4 Self-assessment and subject opinion	115
3.3.5 Inter rater reliability	118
3.3.6 Skills Transfer Results	119
3.4 Discussion	122
3.5 Conclusions	128

Chapter 4 The effectiveness of endovascular simulator training for novices: can learning take place in the absence of expert feedback?

	129
4.1 Overview	130
4.2 Materials and Methods	134
4.2.1 Ethical approval	134
4.2.2 Setting	134
4.2.3 Simulator	134
4.2.3 Participants	134
4.2.4 Study procedures	135
4.2.4.1 Visuospatial testing	136
4.2.4.2 Didactic teaching	136
4.2.4.3 Procedures	137
4.2.4.4 Groups	139
4.2.5 Performance Assessment	140
4.2.5.1 Simulator metrics	140
4.2.5.2 Video-based assessment	141
4.2.6 Statistical plan	142

4.3. Results	143
4.3.1 Demographics	143
4.3.2 Overall results for time, fluoroscopic time and contrast use	144
4.3.3 Between group differences – effect of feedback.	147
4.3.4 Between group differences for video assessment scores	148
4.3.5 Effect of other factors	151
4.3.4 Self –assessment	152
4.3.5 Inter-metric correlation	154
4.4 Discussion	156
4.5 Conclusions	161

Chapter 5 Training surgical skills using non-surgical tasks – can Wii improve surgical performance? **163**

5.1 Background	164
5.2 Materials and Methods	167
5.2.1 Setting	167
5.2.2 Subjects	167
5.2.3 Randomisation	167
5.2.4 Apparatus – Nintendo Wii	167
5.2.5 Study Procedures	154
5.2.5.1 Details of testing session	168
5.2.5.2 Nintendo Wii practice	169
5.2.6 Statistical analysis	172
5.3 Results	172
5.3.1 Results for Bead Transfer Task	173
5.3.2 Results for Glove Cutting task	175
5.3.3 Results for ProMIS virtual task.	176
5.4 Discussion	177
5.5 Conclusions	179

Chapter 6 Retention of surgical skills **180**

6.1 Background	181
----------------	-----

6.2 Methods	184
6.2.1 Endovascular subjects	184
6.2.2 HALC subjects	184
6.3 Results	184
6.3.1 Endovascular results	184
6.3.2 HALC results	189
6.4 Discussion	193
6.5 Conclusions	195

Chapter 7 The benefits of a proficiency-based simulator curriculum for colonoscopy	197
7.1 Background	198
7.2 Materials and methods	200
7.2.1 Aims	201
7.2.2 Setting	201
7.2.3 Simulator	201
7.2.4 Participants	201
7.2.4.1 Subjects	201
7.2.4.2 Patients	202
7.2.4.3 Consultant evaluators	203
7.2.5 Study design and procedures	204
7.2.5.1 Introductory teaching session	205
7.2.5.2 Web based learning	206
7.2.5.3 Randomisation	207
7.2.5.4 Proficiency based simulator training	208
7.2.5.5 Assessment of real colonoscopies	209
7.2.5.6 Statistical Analysis	212
7.3 Results	212
7.3.1 Overall results	212
7.3.2 Pre test results	213
7.3.3 Training results	215
7.3.4 Assessment results	215
7.3.5 Interrater reliability	220

7.4 Discussion	221
7.5 Conclusion	227
Chapter 8 Influence of innate abilities on surgical skills	228
8.1 Background	229
8.2 Methods	233
8.3 Results	233
8.3.1 Endovascular subjects	233
8.3.2 Colonoscopy subjects	235
8.3.3 Wii subjects	235
8.3.4 HALC subjects	238
8.4 Discussion	238
8.5 Conclusion	241
Chapter 9 Conclusions	242
9.1 Overview	243
9.2 Principal findings of the studies	244
9.2.1 Feedback	245
9.2.2 Video game training	246
9.2.3 Skill retention	247
9.2.4 Real-world benefit of a structured simulator training curriculum for endoscopic novices	248
9.2.5 High inter-rater reliability of assessments	248
9.2.6 Influence of innate abilities on surgical performance	248
9.2.7 Skills transfer	249
9.2.8 Self-assessment	250
9.2.9 Feasibility of simulator training	250
9.3 Limitations of the studies	251
9.4 Future uses of the data	252
9.5 Final conclusion	255
References	256
Appendices	281

Appendix Chapter 3	282
Appendix 3.1 Post Training Feedback	282
Appendix 3.2 Tray Error Assessment form	283
Appendix 3.3 Proficiency levels	284
Appendix 3.4 Subject self-assessment form	284
Appendix Chapter 4	285
Appendix 4.1 Standardised advice given to subjects	285
Appendix 4.2 Basic candidate instructions	286
Appendix 4.3 Subject self-assessment sheet	287
Appendix 4.4 Assessment sheet for procedural videos - sample	288
Appendix 4.5 Sample procedural report generated by the simulator	290
Appendix Chapter 7	292
Appendix 7.1 Subject Information sheet	292
Appendix 7.2 Subject Consent form	293
Appendix 7.3 Patient consent form	295
Appendix 7.4 Colonoscopy assessment form	297
Appendix i Chapter abstracts	300
Appendix ii Presentations and publications	307

Acknowledgements

I would like to thank the following for their help and support:

Ms AnnMarie Kennedy
Mr Musallam Al-Akash
Ms Dara O' Keefe
Professor David Bouchier-Hayes
Mr Ciaran McDonnell
Dr Stephen Patchett
Professor Oscar Traynor
Mr Kieran Ryan
Professor Arthur Tanner
Professor Anthony Gallagher

All those who assessed colonoscopies including:

Mr Emmanuel Eguare
Professor Des Winters
Mr Joseph Dignan
Professor Frank Murray
Dr Gavin Harewood
Ms Sarah Brophy

All the trainees who participated in the studies

Thesis Abstract

Improving surgical training – maximising the benefits of surgical simulation

Background

Surgical training is undergoing a period of great change, partly due to the increasingly complex and specialised nature of modern surgical practice. Other factors include working hour restrictions and the need for efficient turnover which restrict training time, and the growing awareness of ethical issues surrounding training on patients. Traditional apprenticeship models of surgical training are increasingly unfeasible and are being replaced by skills lab training, utilising bench models and simulators. A growing body of research supports the value of surgical simulators in surgical training. We aim to investigate ways to improve surgical training, investigating factors which may increase the benefits of simulator-based training programmes.

Methods

We trained a total of 80 surgical trainees and medical students to perform a variety of image-guided procedures on virtual reality-based surgical simulators. The full procedures we selected were colonoscopy, hand-assisted laparoscopic colectomy and endovascular renal artery stenting. In addition, we used basic surgical tasks as part of the training programmes. The 3 full procedures and basic tasks were used during the 4 principle studies from which the data was collected. All our subjects underwent training and assessment. Using these procedures, we investigated the impact of proximate feedback on performance, the value of expert versus non-expert feedback,

transferability and retention of surgical skills, trainee self-assessment, the relationship between innate ability and performance and the impact of simulator-based training curricula on real clinical performance. We also examined the internal consistency of performance metrics and outcomes and assessed inter-rater reliability for our assessments.

Results

Our data demonstrate a steep learning curve during repetition of a simulated procedure, highlighting the value of intensive skills lab training. Standardised feedback reduces error commission, and has less effect on more generalised aspects of performance which improve in the absence of feedback. Error commission is a separate performance metric from the performance assessments which are provided by the simulator. Error scores and more general scores have individual merit but clinically relevant performance metrics should be emphasised during training. Feedback from an expert reduces error commission more than feedback by a non-expert, but has less effect on more generic aspects of performance. However, when errors are objectively assessed, feedback from a non-expert who is familiar with the simulator is equally effective to feedback from an expert. Subject confidence rises after intensive skills training sessions, and self-assessment reflects overall performance improvements, although is less accurate for specific performance markers. There is a wide range of inter-subject variability for accuracy of self-assessment, and this self-assessment should be an integral part of all training curricula. Practice on non-surgical tasks such as video games appears to improve performance although our results did not reach statistical significance, likely due to an insufficient training schedule. Innate ability testing is relevant to surgical skills

performance although the relationship is not yet clearly defined. In a more experienced cohort we found a relationship between specific visuospatial abilities and image-guided procedural performance, and in the intermediate and less experienced cohort, psychomotor ability was more relevant. Perceptual ability appeared to be less important. Attaining proficiency improves skill retention which otherwise partially declines in the absence of reinforcement. Procedure specific error scores showed better retention than instrument handling scores, which may reflect the value of a cognitive curriculum. A proficiency-based progression training programme for colonoscopy was created, utilising many of these factors and this led to an improved clinical performance. We found high correlations between instrument handling and generic performance metrics on all the simulators, but lower correlations between error scores and these metrics. All our assessment systems demonstrated high inter rater reliability.

Conclusion

We have identified many ways to optimise simulator-based surgical training. Simulator-based training is likely to become as integral to surgery as to the aviation industry and it's potential benefits should be maximised. The findings from this research could be incorporated into curriculum design for future training courses. Improving surgical training should ultimately improve patient care.

Chapter 1

Background

Chapter 1 Background

1.1 Brief history of surgery

Ambroise Paré, a 16th century French surgeon stated that there were five reasons to perform surgery: “To eliminate that which is superfluous, restore that which has been dislocated, separate that which has been united, join that which has been divided and repair the defects of nature.”

1.1.1 Surgery in ancient times

Basic surgical procedures have been practised since ancient times, as demonstrated by reports of Burr holes performed as early as 8000 BC (Dimopolos VG et al, 2008) and descriptions of various procedures such as fracture repair and skull injury treatment by Hippocrates. Egyptian carvings depict circumcision procedures dating from 2400 BC, and the Ebers Papyrus (Hallmann-Mikolajczak A, 2004), which is a historical document dating from 1550 BC gives information about the surgical treatment of abscesses and tumours, and how to fix bones. Galen was a surgeon who attended to the gladiators in ancient Rome. He was thus ideally placed to discover more about human anatomy and made huge advances in the field of anatomy (Nutton V, 2007). Despite this evidence of surgical knowledge, the practice of surgery has not improved continuously since these early procedures. In fact the practice of surgery stagnated for many centuries. This was largely due to ignorance about basic sciences such as anatomy, physiology and pathology, and this also affected the practice of medicine. Theories abounded such as that of the four humours. According to Hippocrates, there

were four basic elements that made up the human body. The four humors were identified as black bile, yellow bile, phlegm, and blood. Greeks and Romans, and the later Muslim and Western European medical establishments that adopted and adapted classical medical philosophy, believed that each of these humors would wax and wane in the body, depending on diet and activity. When a patient was suffering from a surplus or imbalance of one fluid, then his or her personality and physical health would be affected (Keirsley D, 1998). This theory was closely related to the theory of the four elements: earth, fire, water and air - earth was predominantly present in the black bile, fire in the yellow bile, water in the phlegm, and all four elements were present in the blood.

A set of characters was developed based on the humors. Those with too much blood were sanguine. Those with too much phlegm were phlegmatic. Those with too much yellow bile were choleric, and those with too much black bile were melancholic. This theory explains one of the most common medicinal practices, that of bloodletting. This was done for a variety of ailments in an attempt to regain equilibrium.

1.1.2 Surgery as a profession

In ancient Greek and Roman times and more recently, the practice of medicine and surgery have been currently closely aligned but for a large part of the Middle Ages, surgery was viewed as a trade which was inferior to the profession of medicine.

Physicians were educated in universities, spoke Latin and were professionals whereas surgical procedures were performed by tradesmen, who performed other tasks such as hair cutting and blood letting. In Paris, physicians in training in the 13th century had to

swear an oath that they would not perform surgical procedures. As far back as the 10th century basic surgical procedures were performed by the clergy, but this practice came to an end in 1215 when the pope forbade the practice of surgery by priests (http://www.barberscompany.org/historical_group.html). Around the 11th to 14th centuries, surgeons existed on a par with barbers and were summoned to perform all manner of tasks such as shaving and teeth pulling and blood letting. (Himmelman L, 2007) This explains why red and white striped poles are still seen outside certain barber shops today, the pole represents the stick grasped by patients to dilate their veins, and the red and white stripes signify blood and bandages. Barber-surgeons were granted guild status in 1308 by King Edward. This arrangement was formalised in 1540 when the Company of Barber Surgeons was formed by King Henry VIII. Many tradesmen were members of similar guilds. Although part of the same guild, the different nature of the two professions was acknowledged in some quarters such as at the College de Saint Come in Paris, probably the first institution where any kind of formalised surgical training took place, where surgeons wore a long robe and barbers a short one. Gradually the more specialised nature of surgery was recognised. Ambrose Pare for example began his career as a common barber-surgeon but then elevated his status by studying medicine. He is known for advancing amputation techniques. Surgery received a boost when King Louis XVI had a fistula successfully repaired and his son Louis XV established academic chairs of surgery at the College de Saint Comme in 1724. Eventually increasing recognition of the importance of surgery led to the dissolution of the barber-surgeons guild, and the two groups were permanently separated in Great Britain in 1745. The Royal College of Surgeons in the UK was formed in 1800.

1.1.3 Important developments in surgery.

A number of important developments towards the 19th and 20th centuries revolutionised the practice of surgery.

Prevention of bleeding

Traditionally blood letting was seen as a therapeutic procedure which could cure a wide range of medical problems. It was standard practice to remove blood before an operation as this was thought to prevent inflammation. Before the era of modern surgery, bleeding to death was a major complication of surgery. Early attempts at preventing bleeding involved the cautery of wounds, which was effective but painful and very destructive of tissues. Whole wounds were burnt, not just the vessels themselves and this frequently occurred when substantial blood loss had already taken place. Such “delicate” substances as boiling oil were used to stem bleeding. The invention of ligatures, by which means blood vessels could be tied off, has been ascribed to Albukaism in the 10th century, but their use was refined by Ambrose Pare in the 16th century. Pare started his career as a common barber- surgeon, and it was during his time treating soldiers on the battlefield that he made advances in the control of bleeding, particularly from gunshot wounds (Mc Callum JE, 2008; Hollingham R, 2008). He later published a treatise detailing his discoveries. Another important development in the field was the work on blood grouping in the 20th century which made it possible to replace blood which had already been lost.

The discovery of anaesthesia.

Prior to this, the ability of the patient to tolerate surgical procedures while conscious meant that undergoing an operation was a terrifying affair and that surgery had to be performed as quickly as possible. Robert Liston, a famous English surgeon in the 1800s was reportedly able to carry out a leg amputation in 25 seconds, although on one occasion he also amputated the patient's testicle and assistant's finger. One of Liston's most famous sayings at the beginning of operations was "Time me Gentlemen!". Although he was not the first to use ether to anaesthetise patients, in 1846 a Boston dentist named William Morton convinced John Warren to use ether during a public operation at Massachusetts General Hospital. (Hollingham R, 2008) The demonstration was a great success, and surgical anaesthesia using ether quickly became a standard in the US and Europe. Anaesthesia gave surgeons the time to operate precisely and refine their craft. John Snow in Edinburgh developed the practice further by using chloroform to anaesthetise his patients. The development of muscle relaxants also facilitated developments in surgical technique.

Antisepsis

The other important development was the introduction of antiseptic techniques. (Hollingham R, 2008) The first surgeon to make advances in the area was Ignaz Semmelweis, a Hungarian doctor who was concerned about the high rates of puerperal fever in his obstetric patients. (Carter KC, 1994) He observed that infection rates were higher in those women attended to by doctors than midwives and speculated it might be due to the fact that doctors at the hospital also carried out autopsies. He introduced hand washing for the doctors before they attended to their patients which dramatically reduced the incidence of puerperal mortality. He did not

fully understand the mechanism for the carrying of infection and surmised it was due to matter transferred from the dead bodies and his theories did not find widespread acceptance during his lifetime. Louis Pasteur was another important figure in the field as he developed germ theory and made many important discoveries about microorganism growth. This led Joseph Lister, a British surgeon to use phenol to prevent infection during surgery. He also began the practice of sterilising surgical equipment and using rubber gloves which dramatically reduced operative infection rates. He published his discoveries in a series of articles in *The Lancet* and his work laid the foundations for antiseptic techniques in modern operating theatres.

1.2 Introduction to endoscopy

1.2.1 Early endoscopy

Endoscopy is a term which describes the visualisation of any cavity of the body with a viewing instrument for diagnosis or therapy. This broad concept dates back to much earlier times in medicine. The earliest description of an endoscopic instrument came from Hippocrates (460-375BC) who described among other things the use of a rectal speculum. Other early speculums were used to view areas of the body such as the cervix and nose and ear. Viewing of more inaccessible body cavities was hampered by the simplistic instrumentation available at the time, and it took many years before technology enabled the development of endoscopy.

1.2.2 Developments in endoscopy

The ability to reflect light into the area of interest was one of the biggest obstacles to overcome in the development of endoscopic surgery. Early attempts used ambient light, but these meant structures were poorly visualised. (Berci G, 2000) One of the first figures to attempt to overcome this was the Arabian Albukaism (980-1037) who improved upon earlier speculums by using reflected light when examining the cervix. The next development came many years later when Giulio Cesare Aranzi (1585) was the first to use a light source for an endoscopic procedure when he reflected light through a flask of water to illuminate structures in the nasal cavity. A significant advance was made in 1806 by Philipp Bozzini, a German doctor and innovator. He invented a light source to be used in endoscopy which although rudimentary was one of the first to use a mirror between the eye and the light source so that the light would be reflected only towards the area of interest and not into the examiner's eye (Berci G, 2000). He used this light guide in combination with various tubes which could be used to examine the mouth, rectum and even the vocal chords. His contemporaries at the time however were sceptical about this new device and dismissed it as a "new toy". It certainly was an important development but the intensity of the light was poor. Bozzini's idea was developed by Antonin J Desormeaux who is often referred to as the "father of cytoscopy" (Ozkara H, 1992) He used an open tube to visualise the urinary bladder in combination with an improved light source. This consisted of a kerosene lamp which burnt alcohol and turpentine to enhance the flame and a lens which directed it to one spot, thereby achieving more concentrated illumination.

The field continued to develop, and the first internal light source was invented in 1867 by Bruck, a dentist from Breslau. The first light sources were cumbersome and even dangerous as they were usually glowing platinum wires which carried a risk of overheating, and causing burn damage to tissues. Things improved slightly with the

invention of the electric light bulb in 1880, which was added to the distal end of the laparoscope in 1883 by Newman. This was still awkward, as there was still a risk of overheating, in addition to poor illumination and colour distortion.

Although some developments were made by the use of separate channels within the endoscopic device, the next important advance came in the 1950's. The technology of fibre optic cables had been discovered in the 1930s. Due to the use of glass of different densities and the physical properties of light, bundles of glass cables could transmit light even around corners without losing energy. This technology was incorporated into a prototype gastroscope in 1955 by Hirschowitz, and the modern era of endoscopy dates from this point. Hirschowitz interestingly demonstrated the use of this new diagnostic tool by performing a gastroscopy on himself.

Along with the developments in illumination came developments in the tubing used in these viewing devices. The earliest devices were simply rigid, open tubes which afforded limited viewing of internal structures. Developments in telescopes and optics led to improved viewing devices. Max Nitze, an urologist from Berlin developed the first cystoscope in 1879 which used electrical lighting and lenses to magnify the object of interest. He collaborated with opticians to make this development. As magnification continued to improve, endoscopists were hampered by the rigid nature of the tube. Semi flexible instruments were developed initially. Parts of these instruments could be angled by a particular amount, such as the instrument developed by Mikulicz in 1881 which could be angled by 30 degrees near its lower third. These instruments worked by using lenses inside the tube to reflect images around bends, but this meant that there were limits on the amount of curvature which could be achieved. Fully flexible instruments relied on a different principle, the use of cameras.

1.23 Cameras in endoscopy

The first endoscopic device to use a camera could more accurately be called a gastrocamera, instead of a gastroscope. It was developed by Lange and Melzing in 1989, and used a small roll of film which was placed at the distal end of the endoscopic instrument and pulled after each exposure. Fifty exposures were taken per examination, providing the endoscopist with a series of images. This design became more advanced, with improvements in the illumination and quality of images obtained. Fibre-optic technology advanced the development of the first gastroscope. In addition to transmitting light as discussed previously, when the individual fibres are arranged coherently, in other words an identical arrangement at either end of the instrument, they can be used to transmit images with no distortion.

The invention of the closed camera device led to the creation of the first video endoscope in 1983, and since then colonoscopy has been one of the most important diagnostic tests for a wide range of gastrointestinal pathologies.

1.24 Modern era of colonoscopy

Colonoscopy is the most accurate means of evaluating the colonic mucosa, and allows biopsy of suspicious lesions. Colonoscopes have flexible shafts that accommodate the curves of the colon, high-resolution optics with magnification at close range, and instrument and suction channels that allow washing, mucosal biopsy, and electrocauterisation. Video imaging and instant printing make a permanent record available. Complete examination of the colon to the cecum can be accomplished in over 95% of patients. The potential discomfort of the procedure is somewhat dependent on the operator, but in most cases the procedure can be comfortably

performed with modest intravenous conscious sedation. Although the risk of colonic perforation or bleeding exists, complication rates for diagnostic procedures are less than 0.5%. The combined diagnostic accuracy is 90-95% for detecting polypoid lesions. Small polyps located behind folds, intramucosal lesions in the setting of ulcerative colitis, and abnormalities present in areas of extensive diverticulosis, stricture, or spasm account for missed lesions. Colonoscopy is approximately 12% more accurate than air contrast barium enema, especially in detecting small lesions such as adenomas (studies claim both lower and higher rates). The cost of the examination is an important issue, especially in reference to the value of screening examinations (see the discussion following). Colonoscopy is most accurate and highly cost effective in the evaluation of symptomatic patients, however.

1.25 Complications of colonoscopy

Colonoscopy is regarded as a safe procedure, and complications are rare. The main complications of colonoscopy are perforation, bleeding, infection and hypotension.

Perforation can be caused by rough handling of the endoscope and pushing against undue resistance, as described above. More commonly, however, it is due to endoscopic snaring of a broad based polyp. Occasionally, perforation can occur because of insufflation of air into a thin walled diverticulum.

If frank perforation into the peritoneal cavity is noted at the time of colonoscopy, then the patient should go for immediate surgical repair. If perforation is suspected, then a water soluble contrast enema should be carried for confirmation. Again, if there is free leakage of contrast into the peritoneal cavity, repair is indicated. Conservative treatment of large colonic perforations is highly dangerous; when perforation occurs

during colonoscopy with a well-prepared bowel little peritoneal contamination occurs and there is a window of opportunity to repair the perforation with minimal morbidity. Taking an expectant approach is likely to lead to abscess formation or generalised peritonitis.

Bleeding is nearly always the result of polypectomy and patients should be warned about this possibility. It may occur at the time of polypectomy or may be delayed for up to 14 days after polypectomy.

Other complications include infarction, hypotension and bradycardia.

1.3 Overview of laparoscopy

1.31 History of laparoscopy

Laparoscopy can be described as a form of closed-cavity endoscopy, and advances in the field of endoscopy as described above were instrumental in the development of laparoscopy although traditionally laparoscopy has only required the use of rigid instruments. Gynaecologists played an important part in the development of laparoscopy, although today its widest application is among general surgeons. The earliest laparoscopies were carried out with a cystoscope and probably the first published report of a laparoscopy was in 1902. This was carried out by George Kelling in Germany, who used a cystoscope to examine the peritoneal cavity and its contents in a living dog using filtered air to create a pneumoperitoneum. (Zucker KA, 2001) He described the operation as a koelioscopic. Kelling repeated the procedure successfully, and also claimed to have carried it out in humans, although Jacobaeus, a

Swedish surgeon published reports of a series of laparoscopies in humans at around the same time and claimed that he was the first to perform the procedure. (Hatzinger et al, 2006) He used a similar technique to Kelling but did not create a pneumoperitoneum. Another pioneer who published results at the same stage was van Ott from Sweden. What he performed was more similar to current day NOTES surgery as although he also inspected the peritoneal cavity, he did this through an incision in the vaginal wall and used a mirror to reflect light and a culdoscope to view the abdominal contents. From these beginnings laparoscopic surgery developed. Many developments in instrumentation occurred over the next few years such as the development of trochars, needles to introduce pneumoperitoneum and the widening of the viewing angle afforded by the laparoscope. Interestingly the Veress needle when first introduced was held to be a safe device as the design mechanism covered the point of the needle after insertion. The inventor Jan Veress however had originally designed it to create a pneumothorax. Although still used by some surgeons and most commonly in gynaecological surgery, it is now recommended to use the open “Hasson” technique for umbilical port insertion as this is safer (Vilos GA et al, 2007)

Although a pneumoperitoneum had been introduced by Kelling when he performed his landmark laparoscopy on a dog, the reason for this was the tamponade effect produced by high insufflation pressures with the aim of reducing bleeding in the stomach or small intestine. The importance of using a pneumoperitoneum was recognised but the optimal gas was not at first. In 1933, Ferrers recommended CO₂ as he had become aware of the potential hazards of using oxygen in combination with a high frequency electrical current inside the abdomen, and reported audible explosions and flashes of light. Another important development which helped to advance

laparoscopy from a purely diagnostic procedure to a therapeutic one was the “dual-puncture technique” which was first used by Kalk in 1929.

1.32 Laparoscopy outside of general surgery

Over a good portion of the 20th century, advances in laparoscopy were made principally by gynaecologists as the technique was largely abandoned by the general surgical community. Improvements in instrumentation facilitated the performance of procedures such as tubal ligation and uterine suspension. Kurt Semm from Kiel was responsible for many innovations such as devices to maintain pneumoperitoneum, safer cautery devices as many injuries were reported with the use of monopolar cautery, laparoscopic instruments such as the cautery hook, atraumatic forceps, and clip appliers. By the 1970’s laparoscopic techniques were being used for procedures such as ectopic pregnancy management, tumour biopsy and staging, and endometrial ablation.

The first laparoscopic appendicectomy was carried out incidentally by Semm in 1982 (Semm K, 1993) during a gynaecologic procedure. From this first procedure the field has now advanced considerably, and laparoscopic techniques have now permeated almost every area of surgery.

1.33 Modern era of laparoscopy

It was not until the era of the laparoscopic cholecystectomy that general surgery began to regain an interest in laparoscopic surgery. There was controversy about who was

first to perform this procedure. Erich Muhe, a German surgeon had been interested for some time in Semm's advances in the field of laparoscopy. In the early 1980's, some internal medical specialists had already performed procedures such as ERCP, endoscopic duodenoscopy and endoscopic papillotomy. Muhe was aware of the need for general surgeons to become more involved in the development of laparoscopic abdominal surgery which had been led by the gynaecological community up to that point. In 1984 he adapted some gynaecological instruments to produce a "galloscope" and in 1985 used this instrument to perform the first laparoscopic cholecystectomy, using the same techniques reported by Semm. (Muhe, 1992) The operation took two hours and went well; afterwards even Muhe himself was fascinated by the rapid recoveries of patients who had surgery performed minimally invasively. He continued to adapt his approach and by 1987 had performed 97 endoscopic cholecystectomies (Muhe, 1992). Despite this impressive output, the response to his work was muted. He presented his work at several meetings and conferences in Germany, but the response from other surgeons was sceptical, and very few contacted him for more information about the technique. At the same time as Muhe's work, similar developments were occurring in France. Philippe Mouret was a French surgeon who shared his practice with a gynaecologist; in 1987 he borrowed some of his colleague's instruments and performed a laparoscopic cholecystectomy. He did not publish this work, but news of his achievement spread to Paris (Schollmeyer T et al, 2007) The first Parisian LC was carried out the following year in 1988, news of these achievements soon spread beyond the borders of France, and some of this work was presented at a SAGES meeting in Louisville in 1990. Many developments had occurred in the application of laparoscopic techniques to this one procedure. The first series of laparoscopic colectomies was reported in 1991 (Jacobs M et al, 1993), and in the intervening years

expertise and technology have improved to the extent that most colorectal procedures are now undertaken using a minimally invasive approach .

As experience and expertise with MIS have grown, the amount of pathology that can be surgically managed with this approach has increased, with ever more complex procedures now successfully performed laparoscopically. Doubts about the oncological efficacy of laparoscopic colon resections for cancer had appeared during the early 1990s due to reports of port-site metastases and earlier recurrence but a number of large randomised, controlled trials conducted in the US, Europe and the UK showed no difference in outcome when comparing open and laparoscopic colon resections (COST group, 2004; COLOR group, 2005; Guillou J et al, 2005) However, MIS was associated with several short-term benefits such as reduced pain, shorter recovery time and reduced post-operative hospital stay. Due to this growing body of evidence MIS techniques for colorectal pathology have become more and advanced, and laparoscopic surgery is now considered the default pathway in many centres.

Although the advantages of laparoscopic surgery are evident, it was noted when first introduced that there were higher rates of particular complications then occurred with open surgery. This was studied by a group from the Royal College of Surgeons in England in 1995 that looked at a large series of laparoscopic cholecystectomies (Dunn D et al, 1994) Post-operative complications such as wound infections, respiratory complications etc were reduced but there was a higher rate of haemorrhage and bile duct injuries which were potentially serious complications. These were shown to be more common during a surgeon's early experience with the procedure, i.e. when they were learning this new technique. Another study was conducted in the US by the Southern Surgeons Club, also examining biliary injuries in laparoscopic

cholecystectomy (The Southern Surgeons' Club, 1995) This group found that biliary injuries were far more likely to occur during a surgeon's first 10 cases, providing further evidence of a steep learning curve when MIS techniques are introduced. A clear learning curve has also been demonstrated for colorectal surgery (Tekkis P, 2005; Bennett CL et al, 1997), and most of the large published trials have a requirement that surgeons must have carried out a particular number of cases before becoming eligible for inclusion in the trial

1.34 Hand-Assisted Laparoscopic Colectomy (HALC)

The first hand-assisted laparoscopic procedures were carried out in the early 1990s during the same period as the first totally laparoscopic colectomies, and the first public presentation on the technique was given at the Euro Surgery meeting in London in 1992 by Patrick Leahy, a colorectal surgeon. Although not an open procedure, in this modification of the laparoscopic approach, the surgeon's hand is inserted into the abdomen to facilitate the dissection. The technique was initially greeted with scepticism; instead of viewing it as an approach which combines the patient benefits of a laparoscopic procedure with the surgeon's advantages of an open procedure, many viewed it as a technique which lacked the benefits of either approach, with some surgeons wondering why only one hand was used instead of two. (Hasson et al, 2007)

In the hand-assisted approach, a laparoscope is inserted and a pneumoperitoneum created, but an incision of between 6 and 12 cm is made, through which the surgeon's hand is inserted. The operation is then performed by using laparoscopic instruments

through standard ports, and the operating hand itself which remains in the abdomen but can be withdrawn and reinserted as necessary. The insertion of the hand solves many of the difficulties associated with laparoscopic surgery. It restores tactile sensation, resolves some of the difficulties caused by the loss of 3 dimensional visions as it facilitate depth perception and makes the operation easier ergonomically. In addition it allows the surgeon to perform blunt finger dissection and facilitates intracorporeal knot tying and haemorrhage control (Darzi A et al, 2001). Cautery, fine dissection and stapling are carried out using laparoscopic instruments as for a standard laparoscopic approach, but even the use of these instruments is made easier as the operating hand can present the target tissue to the appropriate instrument. Therefore the use of the hand has the potential to shorten the learning curve associated with laparoscopic surgery.

One of the criticisms of the hand-assisted approach is that some of the advantages of the laparoscopic resection are lost. The incision which is used to insert the hand is larger than the incisions made for a standard laparoscopic colectomy. However, the specimen extraction incision which is on average x cm in length must be made by the end of the procedure, it therefore seems logical to make use of this incision earlier in the procedure by using a hand port. The hand port must ideally allow insertion and withdrawal of the operating hand without loss of pneumoperitoneum and there are many devices available which can achieve this.

Hand-assisted laparoscopic surgery (HALS) was introduced in an attempt to facilitate the transition from open techniques to minimally invasive procedures. Continuing debate exists about which approach is to be preferred, HALS or totally laparoscopic procedures. Most studies show that the hand-assisted approach is associated with a

shorter operative time and equivalent operative outcomes such as length of hospital stay and blood loss (Ringley C et al, 2007; Aalbers et al, 2008; Boushell et al, 2007; Haasan I et al, 2008) and compared to complete laparoscopy, HALS may be more feasible and offer a lower conversion rate for more technically difficult operations. As such, it is a relevant and valuable operating modality and may be particularly useful as a training operation or intermediate step towards full laparoscopic surgery.

1.4 Overview of endovascular surgery

1.4.1 Overview

Over the past decade, therapeutic options for the treatment of vascular disease have changed greatly. In a similar fashion to the development of minimally invasive techniques for general surgery, the trend is now towards the development and refinement of catheter-based minimally invasive interventions for the treatment of vascular pathology. Endovascular procedures were originally designed for diagnostic use. The basic techniques involve the introduction of a catheter percutaneously into a large blood vessel. Typically the blood vessel chosen is the femoral artery. Access to the femoral artery for example, is required for coronary, carotid, and cerebral angiographic procedures. The catheter is injected with a radio-opaque dye that can be seen on fluoroscopy. As the dye courses through the blood vessels, characteristic images are seen by experienced viewers and can assist in the diagnosis of diseases such as atherosclerosis, vascular trauma, or aneurysms.

In recent years, however, the development of intravascular balloons, stents and coils have allowed for new therapies as alternatives to traditional surgeries such as

Coronary artery bypass surgery (CABG), carotid endarterectomy and aneurysm clipping. Stents and coils are composed of fine wire materials such as platinum that can be inserted through a thin catheter and expanded into a predetermined shape once they are guided into place.

1.4.2 Early developments in endovascular surgery.

Endovascular therapy has continuously evolved since it was first described in 1904. The physicians who paved the way for modern-day endovascular coil occlusion of aneurysms began their work in the late 18th century. Several investigators began experimenting with procedures involving insertion of foreign bodies (needles) into aneurysms to induce thrombosis. (Ransohoff, 1886) Unpredictable results led to the abandonment of these practices. Failures of the early attempts to induce aneurysm thrombosis with percutaneous needle insertion led to a renewed interest in medical interventions, the earliest being potassium iodide for syphilitic aneurysms and aneurysm-related pain, although other substances were also used. These remedies, which were not based on sound scientific evidence, were used with different degrees of success and failed to find widespread application. Better knowledge of the effects and possible applications of electricity led to the development of electro-thrombosis or galvanopuncture (Duncan et al, 1867) The mechanism by which electricity produced thrombus formation was largely speculative. Some of the explanations suggested included inflammation, oxidation, or albumin decomposition,

In 1864, Moore et al postulated that inserting a wire into an aneurysm would provide an ideal environment for clot formation, as opposed to the insertion of a simple

needle, leading to some clinical improvements. In 1879, Corradi attempted to combine wire insertion and electrothrombosis. This procedure came to be known as the Moore Corradi method and spread across North America. Gradually technological advances and improvements advanced the field. In 1964, Luessenhop and Velasquez reported safe catheterization of the internal carotid artery by using silastic tubing. Developments such as this shifted the focus from the extravascular to the endovascular approach. With the introduction of guidewire-supported microcatheters, controlled navigation in the endovascular tree became possible. Further improvements in the endovascular arena took a new direction following the introduction of the Fogarty catheter by T. J. Fogarty and colleagues (Fogarty et al, 1961) This device was developed for the extraction of arterial emboli and thrombi, and led to advancements in the feasibility of balloon catheters. .

1.4.3 Current status of endovascular therapy.

The advantages of endovascular procedures over open vascular procedures for the patient include a less invasive procedure with a shorter recovery time and reduced pain. Studies such as the SAPHIRE trial have shown that carotid artery stenting (CAS) for example is safer than endarterectomy for the management of carotid artery disease in certain high risk patients (Yaday et al, 2004) and the FDA has approved its use in this cohort. Elective endovascular repair of infra renal abdominal aortic aneurysms (EVAR) is one of the more advanced forms of endovascular therapy and has become more commonplace. The technique was first described by Parodi in 1991. Although only certain patients are suitable for this technique, ie with certain restrictions as to aneurysm size and location, early and intermediate outcomes show

reduced operative morbidity and mortality and shorter hospital stays (Greenhalgh et al, 2004; Leadhal et al, 2004) The skill set required to perform these procedures differs from the skills required to perform open vascular surgery. The surgeon must be able to perform an image-guided procedure while using instruments and devices with limited degrees of freedom. It has been shown that there is a learning curve associated with these procedures, such as CAS where a clear correlation has been demonstrated between case numbers and complication rate (Yaday et al, 2004) It is clear that interventionalists must be experienced before attempting these more complex procedures. This is said to be one of the reasons for the results of the EVA-3S trial which also compared CAS and endarterectomy, but has to be stopped early because of poor outcomes in the CAS group. In this study, CAS was performed by relatively inexperienced interventionalists (Mas JL et al, 2006).

Endovascular procedures represent an area of cross specialisation, as they are performed by cardiologists, neurosurgeons and interventional radiologists in addition to vascular surgeons. However, it is important to retain skills for open procedures as there will always be patients, due to factors like emergency presentation, severity of disease or anatomical abnormalities who will not be suitable for an endovascular approach and will require open vascular surgery.

Renal artery stenosis is a narrowing of the renal artery, most often caused by atherosclerosis. Rarer causes include fibromuscular dysplasia, and stenosis after radiation therapy. Renal artery stenosis can impede blood flow to the kidney and can result in hypertension and atrophy of the affected kidney artery, ultimately leading to renal failure if not treated.

Most cases of renal artery stenosis are asymptomatic, and the main problem is high blood pressure that cannot be controlled with medication. Deterioration in renal function may develop if both kidneys are poorly supplied, or when treatment with an ACE inhibitor is initiated. Some patients present with episodes of flash pulmonary oedema. When high-grade renal artery stenosis is documented and blood pressure cannot be controlled with medication, or if renal function deteriorates, renal artery stenosis is often treated invasively. It is most commonly treated by endovascular techniques (i.e. angioplasty with or without stenting). A 2003 meta-analysis found that angioplasty was safe and effective in this context, (Norman AJ et al, 2003) There are ongoing clinical trials to compare medical management and angioplasty with stenting to medical management alone. These include CORAL and ASTRAL, both scheduled to report results in 2010. In addition to endovascular treatment, surgical resection and anastomosis are a rarely used option.

1.5 Surgical Training

1.5.1 Historical Surgical Training

Early surgical training was haphazard and unstructured. It was based on an apprenticeship-type system in that it involved observing or assisting another surgeon for unspecified periods of time. Training was made difficult by the fact that knowledge about the structure and function of the human body was poor up until the 1500s as much of it was based on the work of Galen, who had based his teachings on human anatomy on animal dissection alone (Magnier LM, 1992). The work of

anatomists such as Vesalius did much to further the practice of surgery (Persaud, 1984), and public lectures and demonstrations of human anatomy were performed.

Training was essentially a private arrangement between master and student, with the more experienced surgeon passing on his own personal knowledge (Franzese CB et al, 2007) The affiliation with barbers in the Barber-Surgeon Guilds in the 1500s and 1600s contributed to the view of surgery as a trade rather than a profession but the guild system at least guaranteed some form of training. (Bagwell CE, 2005) In 1505 for example, barber surgeons had to follow certain training guidelines and pass an assessment at the end of their apprenticeship period. The aims of the guild system included that only skilled people were allowed to oversee work, and that the training of young people was properly conducted (Chamberland C, 2009).

In England in 1745, the surgeons split from the barbers to form the Company of Surgeons. In 1800 a Royal Charter was granted and the Royal College of Surgeons in London came into being (later it was renamed to cover all of England - equivalent Colleges exist for Scotland and Ireland as well as many of the old UK colonies (Bagwell CE, 2005). In Ireland, The Royal College of Surgeons in Ireland was granted a Charter on 11th February 1784. The chief provision of this was the institution of the Fellowship which divided Graduates into Licentiates and Fellows. The latter could only be obtained by examination taken a minimum of three years following graduation.

Changes in surgical training can be linked to the changing status of the surgeon in society. During the 17th and 18th centuries in Europe surgeons became dissociated from the Barber guilds and began to gain in economic and political power. (Brunton D, 2004) Previously physicians had been regarded as the academic, educated elite and

surgeons as little more than tradesmen. As the education of surgeons changed it began to resemble more and more the education of physicians. Gradually apprenticeship training was replaced by more formal teaching in hospitals, universities, anatomy demonstrations and public lectures. This system presupposes that there is a body of theoretical knowledge which is necessary to complement proficient practice and that there is sufficient professional competence to standardise this knowledge and present it to students. Both of these elements were probably only present in surgery from the 17th century onwards. While initially surgical training was kept separate from medical, the military need for professionals with both surgical and medical skills led to common educational institutes (McCallum JE, 2008). Medical schools which trained both surgeons and physicians were established in Berlin in 1724 and Vienna in 1791. In France medicine and surgery were officially united in the 1790s when three schools were opened by the state which trained both medicine and surgery. Change was slower in Germany but gradually the range of activities which the barber surgeon was permitted to carry out was restricted. By the 19th century, surgery was recognised as a specialised area within medicine. To become a surgeon, a student had to go to medical school and pass all their medical exams first. It was only after this that they undertook further, more specialised surgical training which was often delivered in the hospital itself.

1.5.2 Halsted and the residency system

Many developments in surgical training took place at the beginning of the 20th century. At this stage Germany was regarded as a centre of excellence in surgical education, due at least partly to the control of medical schools by the German

government and the support of higher institutes of specialised learning (O'Shea J, 2008). Since unification in 1870, all proprietary medical school in smaller states had been eliminated by the central German government. The German system was founded on 3 main principles; – knowledge of the basic sciences, the pursuit of research and graduated clinical responsibility (Sealy WC., 1999).

One of the major figures in surgical training is undoubtedly William Halsted. Halsted was a prolific US surgeon who practiced around the turn of the century. He is remembered for many achievements in the clinical practice of surgery and but also initiated a structured surgical training programme which laid the foundations for the current American residency programme (Magnier LM, 1992; Sealy WC 1999). As a trainee he studied in Austria and Germany from 1878 to 1880. Studying in Europe was a common practice for US trainees during that period, and Germany in particular had a reputation as a centre of excellence in medical training. Halsted's European experience is credited as the inspiration for the system of surgical training he introduced at Johns Hopkins in 1889. This system of progressive responsibility involved the trainee surgeon gradually being allowed more operative exposure and less supervision over a set period of time, ending with a period of independent activity. Trainees learnt through observing and assisting attending surgeons, and through prolonged patient contact. The system entailed a very close relationship between trainee and consultant surgeon, and involved such long hours for the trainees that they literally resided in the hospital, hence the name resident. According to an address Halsted gave in 1904 at Yale, it was designed to produce surgeons "of the highest type, who will stimulate the finest youths of their country to study surgery, and to devote their energies and their lives to raising the standards of surgical science." (Halsted WS, 1904) William Osler, who practiced at Johns Hopkins at a

similar time, was an eminent physician and teacher who also was an important figure in the development of the residency system. Osler insisted that his medical students get to the bedside early in their training; by their third year they were taking patient histories, performing physicals and doing lab tests examining secretions, blood and excreta instead of sitting in a lecture hall, dutifully taking notes. He diminished the role of didactic lectures and once said he hoped his tombstone would say only, "He brought medical students into the wards for bedside teaching."

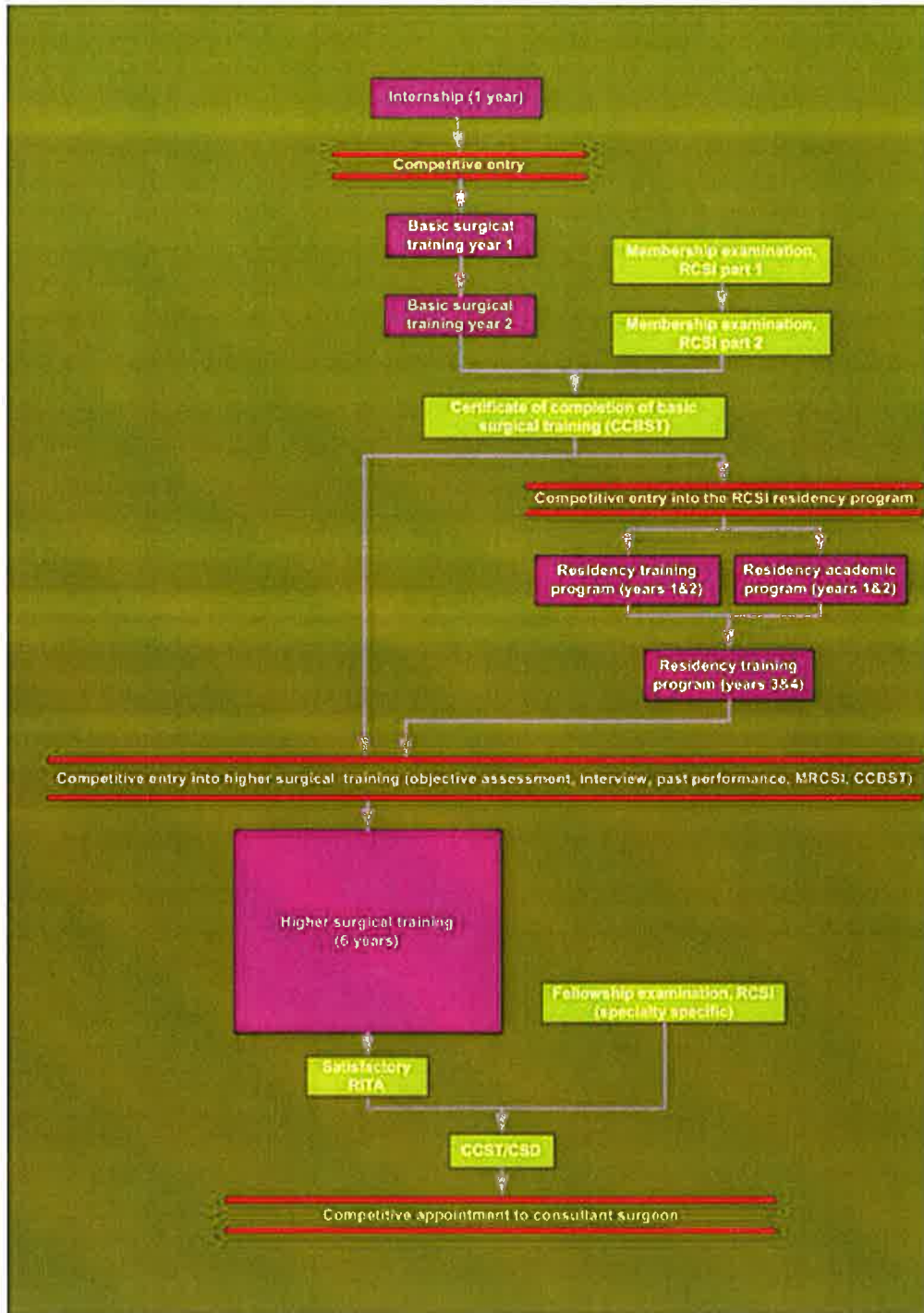
This residency system was a huge improvement on the unstructured, haphazard training which had been available previously, and it laid the foundation for the US residency program, which is still in place today. However it had some deficiencies; it made huge demands on the trainee surgeons, it had an indefinite length, was characterised by intense competition, and it was often known as a "pyramidal" system because there were a large number of house officers entering the scheme, but only one chief resident (Pories WJ, 1999; Sealy WC, 1999).

The Halstedian model even as practiced today demands long hours and is centre- and supervisor-dependent. It is largely governed by the random admission of patients, with little opportunity to standardise a curriculum, or ensure that residents get the same exposure as their counterparts at other hospitals. Teaching is delivered in the operating room, which is not an ideal learning environment for the trainee (or the patient), assessment of the trainee is subjective and they progress when their consultant feels they are ready. Although this training system is long established there are a number of areas within it which need to be improved, and a number of factors which are driving this change. As the demands made on surgeons have changed and

developed over the last number of years, this training model has become increasingly inadequate

1.5.3 Current Irish structure in surgical training

Figure 1.1 - Below is an overview of the current training route from graduation from medical school to consultancy. Competitive entry is highlighted in red, years in training are highlighted in blue and examinations and other requirements are highlighted in green.. Figure is courtesy of RCSI.



Although medical schools teach both medicine and surgery along with other subjects, surgical training formally begins at postgraduate level, after completing medical school and one year of internship. Recognised surgical training programmes in Ireland fall under the remit of the Royal College of Surgeons in Ireland. The Irish model is

similar to programmes in the UK and Australia and can take up to 12 years to complete. There are many more applicants for basic than for higher surgical training as many trainees leave during basic surgical training. (BST) Trainees compete for entry to basic and higher surgical training schemes, and the period in-between may be spent on the residency programme, or can be regarded as the “gap” years when trainees undertake research in order to gain a higher degree and/or publications. This makes them more competitive for selection onto a higher surgical training programme as there is intense competition for these places.

1.5.4 Current issues in surgical training

1.5.4.1 Work hour limitations

Surgical trainees today have limitations on the amount of hours they can spend working. Incidences of fatigue as a cause of medical error have been widely reported and this is one of the reasons for increasing regulation in this area. There is a large body of laboratory data showing beyond a doubt that fatigue impairs human performance (Van Dongen HPA et al, 2000; Dinges DF et al, 1997) In fact, the effect of sleep deprivation on a task that involves tracking has been shown to be equivalent to the effect of alcohol intoxication; in one study, performance of such a task after 24 hours of sustained wakefulness was equivalent to the performance with a blood alcohol concentration of 0.10 percent. (Dawson D et al, 1997) Studies of simulated driving have had similar results. (Arndt JT et al, 2001) Over the past 30 years, many studies have provided unequivocal evidence that mood is worsened by fatigue, as indicated by increased scores on measures of depression, anxiety, confusion, and

anger, and that psychomotor performance is impaired in sleep-deprived residents. (Asken MJ et al, 1983; Samkoff JS et al 1991, Leung L et al, 1992, Owens JA et al 2001)

Studies in sleep laboratories show that both at base line and after on-call duty, levels of daytime sleepiness in residents are similar to or higher than those in patients with narcolepsy or sleep apnoea.

It has been more difficult to prove that sleep deprivation impairs clinical performance. Most, but not all, studies show impaired performance of clinically relevant, although artificial, tasks. (Asken MJ et al, 1983; Samkoff JS et al 1991, Leung L et al, 1992) For example, sleep deprivation affected hand–eye coordination in surgeons performing laparoscopy but did not impair the performance of surgical residents on written board examinations.

The issue of regulating junior doctors' working hours goes back to the 1980s. After problems in New York City, a commission chaired by Dr Bertrand Bell, professor of medicine at Albert Einstein College of Medicine, recommended changes. New York State Department of Health Code, Section 405, also known as the Libby Zion law, is a regulation that limits the amount of resident physicians' work in New York State hospitals to roughly 80 hours per week. (Strongwater 1983). The law was named after Libby Zion, a young woman who died at the age of 18 under the care of overworked resident physicians and intern physicians in a New York hospital. She died within 24 hours of her admission, most likely due to serotonin syndrome caused by an interaction between medications she was already taking and ones which were prescribed on admission. This was never properly diagnosed by the two doctors who treated her on admission, and was blamed on a number of factors including the

excessive workloads the doctors were dealing with at the time and the lack of senior supervision. One of the doctors for example was an intern 8 months out of medical school who was responsible for the care of 40 patients that night. Her parents were unsatisfied with the reasons given for their daughter's death and questioned the competency of the doctors. The lawsuits and regulatory investigations following Zion's death, and their implications for working conditions and supervision of interns and residents were highly publicised in both lay media and medical journals (Spritz N, 1991)

In July 2003 the Accreditation Council for Graduate Medical Education (ACGME) adopted similar regulations for all accredited medical training institutions in the United States. From 1 July 2003, the ACGME limited the work of the United States' almost 100 000 resident physicians in 7800 programmes to 80 hours a week. Shifts do not last more than 24 hours, and residents have one day off in seven and get a 10 hour break between being on call and working a shift. However, many junior doctors do not comply with these restrictions. (Josefson, 1998)

Similar restrictions on doctor's working hours have been applied in Europe. In 1993 the Calman report set out to revise specialist medical training (The Calman Report London: Department of Health 1993) In addition to recommending structured training programmes, progression through training based on formal annual assessments of competence, and much shorter training in most specialties. It limits hours worked each week by junior doctors. In Ireland the future implementation of the European Working Time Directive (EWTD) will impact on the working week for medical and surgical trainees.

The original directive on working time became law in 1993, but doctors in training were excluded, along with workers in the road, air, rail, sea, and inland waterway industries. The British government of the time challenged the validity of the directive as health and safety law, but it was confirmed in 1996 as such by the European Court of Justice.² Although the EWTD should have been introduced for doctors in training as far back as 2004, most rotas in Ireland remained unchanged. The basic rules of the European Working Time Directive (EWTD) are simple: “An adult worker is entitled to a rest period of not less than eleven consecutive hours in each 24-hour period during which he works for his employer” and the average working week for a junior doctor should be no more than 56 hours, decreasing to 48 hours by 1st August 2009; all must have two rest periods of at least 24 hours in every fortnight, and four weeks of annual leave (Department of Health and Children, 2004) Despite the obvious fact that long working hours and fatigue impair performance, there are a number of problems with the restrictions on working hours. Apart from patient related factors such as non-continuity of care, surgical trainees are trained on the job, and reduction in working hours also reduces training hours. There are concerns that capping hours worked will not provide sufficient exposure to surgical trainees. Shift work compounds the problem as most training takes place within the context of the normal working day. A trainee on a EWTD compliant shift for example may have to work a week of nights, during which time there may be little or no consultant contact and little operative exposure.

In addition there is a gradual change in the mindset of surgeons who are now acknowledging the importance of lifestyle and family in addition to their professional life, and this trend looks set to continue given the large amount of female trainees who

are choosing to work in surgery (Antiroll DA, 2007) Reductions in available training hours make the Halstedian training model increasingly unworkable.

1.5.4.2 Increasing complexity in surgery

The changes in surgical practice which the last few decades have witnessed have also driven the reassessment of the adequacy of surgical training. Technological development and higher standards of skill and dexterity are advancing the range of complex procedures which are now carried out. Improvements in intensive care medicine and anaesthesiology mean that older and sicker patients can be operated on than were previously. Surgical training must adapt to prepare its trainees for these challenges. One of the most significant innovations of the last century has been the advent of laparoscopic surgery (Mercy, 2007) ^{and} the expansion in the range of surgical conditions that minimally invasive techniques can be applied to. Although both open and laparoscopic surgery demands high levels of manual dexterity, it is intuitive that minimally invasive surgery (MIS) requires a different skill set. These advanced skills include the ability to reconstruct 3D imagery from a 2D visual monitor display with a loss of binocular depth cues, manipulation of longer surgical instruments, coping with a huge reduction in the amount and quality of tactile feedback and adaptation to the fulcrum effect, or the counter intuitive movement of the surgical instruments due to the function of the body wall as a fulcrum.(Gallagher et al, 2003) When laparoscopic surgery was first introduced for procedures which were normally carried out in an open fashion, complication rates were unacceptably high until a surgeon had performed the procedure a number of times. This learning curve has been extensively studied and can last for between 10 and 60 cases depending on the procedure. (Tekkis at al, 2005) One of the first procedures that

minimally invasive techniques were applied to was the laparoscopic cholecystectomy. For example, a study by The Southern Surgeon's Club in 1991 (The Southern Surgeon's Club, 1991) estimated that a surgeon would have to perform 13 laparoscopic cholecystectomies before their rate of common bile duct injury fell from 2.2% to an acceptable limit of 0.1%. Another study carried out in the UK demonstrated that although systemic complications were lower for laparoscopic cholecystectomy compared to open, rates of bile duct injury were significantly higher. (Dunn et al, 1994) Although the laparoscopic cholecystectomy learning curve is now no longer an issue as surgical trainees perform many of these cases during the course of their training, this research demonstrated that the existing training methods did not train surgeons adequately to perform minimally invasive surgery, and also highlighted the possibility of severe complication rates if a surgeon's first attempts at laparoscopic surgery were in the operating room itself during a live case, as would be the case in the apprenticeship model of teaching, albeit under supervision. The learning curve for other minimally invasive procedures is even longer, for example for laparoscopic colectomy, a surgeon must perform 50 cases before the learning curve starts to level out.

1.5.4.3 Economic factors

Operating theatre time is also at a premium. Economic pressures and long waiting lists mean that high patient turnover in theatre is essential. Training opportunities for the junior surgeon are often missed due to time restrictions. Taking a surgical trainee through a procedure understandably takes longer than it would take for a consultant to perform the procedure themselves, and trainee operating times have been shown to be

longer than consultant operating times in a number of studies, (Lutz et al 2009)

During the last decade, as reimbursement for services provided by hospitals and physicians has declined and as technologic advances have provided more sophisticated (and expensive) ways to provide care, hospitals have become more concerned about their cost structure and the inefficiencies that result from the learning curve associated with the apprenticeship model (Babineau et al. 2004, Bridges et al 1999, Gold et al 2004)

Long waiting lists and increasing privatisation of healthcare have also directed many elective surgical procedures towards the private sector, meaning that surgical trainees who train at public hospitals get less exposure to elective procedures. Some of these are relatively minor procedures such as varicose vein surgery that trainees would normally be able to carry out independently, so the loss of these procedures to the private sector robs trainees of valuable operating experience.

1.5.4.4 Ethical issues

Ethical issues have also prompted the re-evaluation of apprentice-ship teaching. Medical errors and liability have been more prominent in the public consciousness due to high-profile cases of medical error, such as the Bristol Case (The Senate of Surgery 1997) in Britain, the Neary case (The Lourdes Hospital Inquiry, 2006) in Ireland and the Libby Zion case in the US (Spritz N, 1991). Medical ethics and patient autonomy are now very topical issues. A new structure for residency training must take this change into consideration. It has become less and less acceptable in the current climate to train or practice on patients, and this has reduced learning

opportunities for the trainee. Apprenticeship as the core of residency training was appropriate and socially acceptable during the first part of the 20th century, when patients who had no ability to pay were grateful for the care provided by individuals who were learning. After the establishment in 1965 of Medicare, which gave all Americans over the age of 65 the right to access medical care in a more or less private setting, and the consequent abolition of ward or city-county charity services, society became much more concerned about the use of unsupervised trainees to provide medical care. As principles of ethics evolved in medicine, many have started to question the very essence of providing care while training. (Dutta et al, 2003)

Trainees in surgery start to perform simple parts of procedures under close supervision and are gradually allowed to perform more steps and eventually to operate independently. However at the beginning of their training they may be performing a procedure on a patient for the first time. In addition, trainees are not currently required to attain a certain, objectively defined level of competence before being permitted to perform more of procedure; this a decision made by their supervisor. It has been shown that surgical outcomes are better when trained surgeons operate. Patients have greater autonomy than in the past and some patient may refuse to let a trainee perform any part of their surgery. There are of course other ways to train practical skills but all have limitations.

1.5.4.5 Lack of standards

Another aspect of the ethical problems surrounding surgical training is the lack of standards. Trainees who complete the duration of their training, ie a 6 –year specialist registrar post in general surgery are generally deemed trained or “signed off” despite

the fact that they do not have to display a particular level of competence. Time spent or number of procedures performed are still frequently used as markers of proficiency despite the fact that they do not guarantee a minimum standard of operative ability. While some trainees may indeed become competent after performing for example 50 colonoscopies, and some may become competent sooner, some trainees may require a greater number of procedures to reach competency. Attempts are being made to address this problem, for example in the UK, an endoscopist who wishes to carry out screening colonoscopies as part of the national colorectal cancer screening programme must objectively demonstrate procedural competence. The To Err Human report which was published in 1999 (Kohn LT, 1999) was a ground breaking study of errors in healthcare. The report was based upon analysis of multiple studies by a variety of organizations and concluded that between 44,000 to 98,000 people die each year as a result of preventable medical errors. For comparison, fewer than 50,000 people died of Alzheimer's disease and 17,000 died of illicit drug use in the same year. The report addressed some of the issues surrounding medical and surgical training and one of the recommendations from this report was the introduction and standardisation of competency testing for health professionals

Surgical training has also suffered from being centre and supervisor-dependent. A trainee who works with a consultant who is an enthusiastic teacher in a quiet hospital may have a very different training experience than his peer who is rotating through a large tertiary referral centre on a busy service. Training is over-reliant on the random admission of patients with no way to guarantee that a trainee will have the opportunity to perform a certain index procedure a set number of times.

A recent issue in the provision of healthcare has been that of centralisation of specialist services, particularly centralisation of cancer services. This is due to evidence showing the influence of hospital or surgeon case-loads or specialisation on the survival and quality of care of cancer patients. The process is currently underway in Ireland, with the selection of 8 major hospitals as cancer centres. This is likely to impact on surgical trainees as a trainee who is rotating through a non-specialist centre will have less exposure to certain procedures, and is another reason for training inequalities.

1.5.4.6 Competency

The goal of any surgical training programme is to produce surgeons who can competently perform procedures, but precisely what constitutes competent is difficult to define. Traditionally in surgical training, measures of learning such as amount of cases performed or time spent training have been used. Similar measures are used in other disciplines such as aviation training where a certain number of flight hours must be logged by training pilots. Although measures such as these do give some indication of a trainee's standard they are at best crude indicators of ability as individuals learn at different rates. Some may become competent in a task or procedure after only a few attempts, others may require many more trial runs. In addition there are other factors which govern rates of learning such as initial experience level, motivation and fundamental abilities (Gallagher AG et al, 2005).

As far back as the 1980s, the American College of Physicians appointed specialist committees to make recommendations about standards which must be attained for

credentialing (Roberts JS et al, 1988). As a result, many different disciplines set a minimum number of procedures which had to be performed by a trainee before they could be deemed competent in that procedure, such as in colonoscopy, where 140 procedures (Cass OW, 1996) was recommended by the ASGE, although this was specified as a rough benchmark level (Principles of Training in Gastrointestinal Endoscopy 1999), which might vary for different individuals. However although this was another attempt to standardise competency levels for proceduralists and was convenient to use, it still consisted of numbers of cases performed, and as such does not guarantee proficiency. Other problems with this system are that many different credentialing bodies require different standards for the same procedure and that once a minimum number of procedures has been specified it often becomes an accepted benchmark for every trainee (Wigton RS 1999). In addition it has been shown that many trainees need to perform more procedures than the minimum required to feel comfortable with that procedure (Wiggs CM et al, 2000).

Proficiency-based progression is a simple concept that overcomes many of these problems. It is a system whereby a trainee is only deemed competent in a particular procedure if they can demonstrate that they can perform the procedure to a certain predefined standard. They can only progress to the next, more complex stage or task when they have attained proficiency in the current task. Although relatively recently applied to surgical training, the concept is well established in other areas. In the US, the Association for Continuing Graduate Medical Education (ACGME) has published a set of 6 core competencies which must be objectively assessed at individual centres to ensure that trainees are meeting minimum standards.

Although surrogates of competency such as logbook data and number of years spent training are still being used to assess trainee ability, it is likely that surgical training and assessment will continue to move towards a competency-based approach.

1.5.4.7 Feminisation of surgery

Although approximately half of entering medical students are female, only 6.5 % of consultant surgeons are women (RCSEng). Women have traditionally been less inclined toward a surgical career (Noviellei et al, 2001) The reasons for this are not clearly understood. While lifestyle and family issues may have been assumed to be important, there is evidence that the actual reasons are more complex and some deterrents to a surgical career might be relatively unique to women (Gargiulo at al, 2006). It has been shown in one study that both men and women are deterred from a career in surgery primarily by lifestyle considerations. However, the surgical culture and personality are sex-specific deterrents to women. (Gargiulo at al, 2006). These factors create substantial challenges to surgical educators (Polk et al, 1999). Whatever the reason, this under-representation carries serious implications for the future of the profession. If surgery continues to fail to attract women entrants, the surgeons of the future will have to come from a declining number of male medical graduates. It is important to attract those with the highest aptitude for surgery, regardless of gender. Certain training programmes have introduced more flexible training hours and job sharing in a bid to attract more female trainees. The integration of simulation into surgical training programmes has the potential to shorten training hours and introduce greater flexibility, thus removing some barriers to female surgical trainees.

1.5.5 Training platforms

The need to gain technical skills outside of the clinical arena has long been recognised and has led to the development of many alternatives, from simple bench models to complex surgical simulators.

Bench Models

Basic bench models are simple but effective trainers. Even simple, everyday materials can be used to train the techniques needed for more complex procedures. A historical example can be found by examining the career of Alexis Carrell. He was a French surgeon who practised at the Rockefeller Institute and made pioneering developments in the field of vascular anastomosis, which greatly advanced vascular surgery and organ transplantation. He reportedly gained some of his fine manual skills from his experience with complex embroidery techniques involved with making lace (Langer et al, 2002) His mother had worked as a dressmaker. Even simple low fidelity bench models have been shown to be effective training tools (Ravi et al 2007, Matsumoto et al 200?). Matsumoto et al demonstrated skills transfer with the use of basic dental rolls. However in modern skills labs most bench models, while simple consist of materials or models specifically designed to simulate a particular task. For example, skin pads are available made of foam and rubber with different features such as skin lesions or cysts so the trainee can practice removing the lesion and suturing the skin. These models are designed to be used with real surgical instruments, so real scalpels and suture materials are used. What is more important is how the tools are

used in the presence of a structured curriculum. Medical students are frequently instructed to practice their suturing skills on orange peel before attempting to repair a wound on a patient.

Cadaveric Dissection

Other training platforms include cadaveric models, which are a valuable training resource and have been used in medical and surgical education since Renaissance times (Persaud, 1984). Cadaver training gives students an important three-dimensional view of human anatomy which is difficult to gain with textbooks alone. It introduces them to the range of variability encountered in living tissue, allows them to practice dissection skills and integrates specific anatomy with the body as a whole.

Historically shortages of cadavers led to the procurement of cadavers from less than reputable sources, and so called “Renaissance men” in the 1700s and 1800s often supplied dissection rooms in medical schools. Enthusiastic students frequently sourced their own cadavers, throwing the medical profession into some disrepute and fuelling accusations of grave robbing. In the UK, the “Murder Act” of 1752 meant that the corpses of executed criminals were available for dissection, but demand far outweighed supply. The Burke and Hare scandal of 1829 (in which two Irish Immigrants, William Burke and William Hare murdered up to 17 people and sold their corpses to Edinburgh Medical College for dissection) probably led to the Anatomy Act of 1832 which provided for the needs of physicians, surgeons and students by giving them legal access to corpses that were unclaimed after death, in particular those who died in prison or the workhouse. Further, a person could donate their next of kin’s corpse in exchange for burial at the expense of the donee. The Act

provided that anyone intending to practice anatomy must obtain a license from the Home Secretary. One or two teachers in each institution took out this license and were known as licensed teachers. They accepted the whole responsibility for the proper treatment of all bodies dissected in the building for which their license was granted. Regulating these licensed teachers, and receiving constant reports from them, were four inspectors of anatomy, one each for England, Scotland, Ireland and London, who reported to the Home Secretary and knew the whereabouts of every body being dissected.

Although supply of cadavers to medical schools today comes from voluntary donations from members of the public, and takes place in a structured and ethical format, cadavers are available in increasingly limited quantities. In addition they are expensive, difficult to store and have may have different properties to living tissue (MacLaughnlin, 2004). There is also some debate regarding current methods of obtaining consent for cadaveric donation, and suggestions that information given prior to obtaining consent is insufficient (Chung et al, 2002). Other disadvantages include the remote but unavoidable risk of disease transmission (Wiwanitkit V, 2002) and the difficulties and expense involved with treating and preparing cadavers for dissection. For this reason, some medical schools have introduced programmes which do not involve cadaveric dissection. The first UK school to do so commenced its programme in 2002. Alternatives include prosection which is the dissection of a cadaver or part of a cadaver by a professional in order to demonstrate anatomic structure to students. In a dissection, students learn by doing; in a prosection, students learn by watching. This dramatically reduces the number of cadavers which are usually required by medical schools, and although the students do not have the tactile experience of performing the dissection, they will probably have the benefit of a more thorough and

exact dissection with all relevant structures exposed as it has been performed by the instructor. It has been argued that this method of teaching anatomy is as effective as dissection (Bernard, 2005). A prosection may also refer to the dissected cadaver or cadaver part which is then reassembled and provided to students for review. In practice, cadaveric dissection falls more under the remit of medical student education and currently is rarely used for postgraduate surgical training.

Animal models

Living animal models are high fidelity in some aspects and importantly can allow the training of complete procedures, such as the performance of laparoscopic cholecystectomy on pigs (Bailey, 1991). One of the biggest challenges facing the laparoscopic surgeon is the identification and control of bleeding within the pneumoperitoneum, and a live animal model provides a realistic simulation of this situation that would be difficult to achieve by any other means. In this instance pigs are humanely anaesthetised, and then trainees can perform complete, minimally invasive procedures on living tissue, with a beating heart, real bleeding etc. Many US medical schools incorporate live animal labs into their training programmes, and in Europe they are commonly used in the context of short, intensive operative surgery courses. However, such courses are expensive and specialised. A trainee may have the opportunity to perform a procedure once only, and although this is beneficial, it does not allow them to practice until they reach a level of proficiency. In addition it has been demonstrated that the massed learning which takes place during such intensive courses is inferior to distributed practice. The provision of such labs and facilities is expensive; there is currently no live animal training facility in Ireland for example. In addition to the skills lab itself, it is necessary to provide appropriate holding facilities,

food and water, and measures to ensure minimisation of pain and distress, and euthanasia. This would also require specialised personnel. Irish trainees who wish to participate in live animal courses must travel abroad. In order to be a useful training method, trainees should have regular access to such facilities. In addition, ethical considerations have further limited the use of live animals for surgical training. In the UK for example, animals can be used for testing pharmacological compounds but for training purposes. In Australia, live animals can be used in surgical training, but it is stipulated by the Animal Ethics Committee that training methods which do not need animals must be used wherever possible, ie “Training in basic surgical skills such as suturing, knot tying and the anastomosis of hollow viscera and blood vessels can, to a large extent, be performed on non-living material, and most other skills can be learnt directly from more experienced surgeons.”

Virtual Reality simulation will be discussed in the next section.

1.6 Simulation

A simulator can be defined as a model or system which represents the workings or mechanism of another system, as distinct from an emulator which is a system which performs in exactly the same way as another system. Simulators therefore do not need to produce the same output as the system they represent, but rather to behave in the same way, or more specifically a surgical simulator does should not have to look like real surgery, as long as it allows the trainee to practice a procedure realistically. Simulators are perhaps best known as training aids for those who must perform

complex or high risk procedures. They can mimic the procedure or situation which is high risk and allow the trainee to practice in a risk free environment.

1.6.1 Simulation in the aviation industry

The surgical community has had the advantage of being able to witness another area in which simulation has become a central part as simulation has been used in the aviation industry since 1910. Charting the development of simulation in the aviation industry draws interesting parallels with the world of surgery. The initial attempts at simulation in the aviation industry were made without the benefits of today's engineering and computer advances. Even before synthetic flight machines were available, the importance of practising piloting skills before actually becoming airborne was recognised. Early aviation students were allowed to drive airplanes along the ground after passenger flights to practice rudder control, and then they could graduate to making short hops. (Turner LWF, 1913) Planes were even designed for this purpose with reduced wingspans, and this training method was known as the penguin system. The next development was essentially a small plane mounted on a joint which could more closely simulate flying called the Sanders teacher. (Howard DM, 1913) Over the next few years more training devices were built and although all were limited by rudimentary technology and resources, their importance in pilot training was recognised. One of the best known early simulators was the Links trainer, invented by Edwin Link in 1929. This simulator also known as the blue box consisted of a wooden fuselage mounted on an air bellows, which was able to represent the movements involved in flight. Further modifications were added so that it could simulate instrument flying. This allowed a pilot to train for hours before ever going near a plane, and although initially dismissed by some as a fairground attraction or a

toy, the Links trainer would soon become an integral part of flight training. After a number of fatal aviation accidents, the US Air Force purchased 6 Links trainers in 1934 in recognition of the fact that their existing training programme was inadequate. This was an important step in the development of aviation simulation, and soon many other countries purchased Links trainers.

The quality and complexity of aircraft simulators has been steadily progressing over the last 60 years, driven by increasing demands on pilots, such as the need to become proficient in instrument or blind flying, the high demand for military pilots during World War 11 and the development of commercial, large aircraft. By 1955, aviation simulation had progressed to the degree that it was validated by the Federal Aviation Authority and became an obligatory part of annual flight certification. Now all pilots must complete a required number of hours in a simulator before they can fly a plane, and they must train and be certified in a simulator which is specific to the type of aircraft they will fly. Therefore there is no time wasted inside the actual cockpit. In addition to this economic benefit the more important issue of passenger safety has been one of the driving influences in aviation simulation. Passengers would not contemplate boarding a plane which is being piloted by a trainee, and yet every day in teaching hospitals trainee surgeons perform surgical procedures, although they are supervised. Space travel has also driven the development of complex simulators which are used to train astronauts. This allows the astronauts to experience conditions such as weightlessness which they would otherwise have no chance to experience (Ball CG, 2008).

The surgical community is several decades behind the aviation industry as the first surgical simulator was built in 1989, but developments in simulation have been

occurring rapidly over the past two decades, with most modern day simulators incorporating virtual reality technology.

1.6.2 Virtual Reality Simulation

Virtual reality (VR) is a term that has been attributed to Jaron Lanier, the philosopher and computer scientist. He used it in the 1980's to describe the concept of a virtual world which supported interaction, instead of something which could only be passively visualised. The idea of creating virtual environments had been developed over the previous years. What is widely considered to be the first head mounted display system was created in 1968 by Ivan Sutherland. This was a primitive and awkward device which was worn by the user and allowed them to see virtual surroundings which were represented by very simple graphics. Other later systems provided more realistic visual experiences, but the ability to interact with and effect changes in a virtual environment makes this technology a useful tool in surgical training. Virtual reality technology is used extensively in the gaming industry, and has been the subject of many books and films where it is portrayed as a very advanced technology, but it is increasingly being used for training and therapeutic demands

1.6.3 VR simulation in surgery

The first surgical simulator to use VR technology was a representation of the lower limb which was developed in 1989. (Delp SL, 1990). This was a tool to help orthopaedic surgeons plan the optimum location for tendon transplants, as it showed them how the virtual leg would function after a tendon had been transferred to various locations. The next step was probably the simulator which was designed by Jaron

Lanvier and Richard Satava, a pioneer of information technology applied to medicine (Satava RM, 1993). This highlights the optimum way of designing a simulator, ie to have professionals from different disciplines involved from the very beginning. Their simulator represented the organs of the upper abdomen. Although graphics were basic, it permitted the user to view the organs from different angles, and some interactivity facilitated training of the steps required for a cholecystectomy. Simulators for surgery have now become more advanced due partly to improvements in computer processing. Now, graphics are much more realistic, and simulated organs can be deformed in real time and manipulated by simulated surgical instruments.

1.6.4 Developments in surgical simulators

Many simulators now allow for the input of actual human data, via CT or MRI imaging, allowing for the creation of “personalised” simulators which can be specific to a patient or to a pathology. This concept was first introduced in 1989 as the Visible Human Project from the National Library of Medicine. This project involved combining CT, MRI and cryosection images from a real male and female to create detailed visual data for the real human body. The CT axial images for the female for example were obtained at 0.33mm intervals. This allowed 3d representations of any part of the human anatomy to be reconstructed and visualised, and in addition to its importance for medical education had huge relevance to simulation and intraoperative navigation. Data from the visible human project could be applied under licence to other projects such as simulators. Other developments in surgical simulation include the incorporation of real patient data into the simulator using imaging such as CT or MRI. This permits “mission rehearsal”, in addition to training the general skills necessary for the operation, the surgeon can plan and practice an individual patient’s

procedure. This concept was first developed by Jeff Levey in 1993, who utilized it in a hysteroscopy simulator (Levy JL et al, 1993)

Another important capability which is essential in simulation is tracking, or assessment. One of the major advantages of computer based simulation is that the same experience or sequence of events can be replicated repeatedly, which allows the trainee to learn from mistakes in a safe environment. Another benefit which is probably equally if not more important is the objective feed-back a trainee receives from a computer based simulator. Since everything a trainee “does” on a computer based simulator is essentially data, all actions can be tracked by the computer. In addition to crude measures such as performance time, detailed data such as instrument path length, speed of instrument movement, and the exact location in space of any instrument at any point in time is recorded. These data can be used to create a set of very robust and objective performance metrics. A simulator without metrics is really no better than an expensive video game. While the main function of metrics is to provide the trainee with objective and proximate feedback on their performance, they also allow the trainer to objectively assess the progress of the trainee throughout the training process. This allows the trainer to provide formative feedback to aid the trainee in acquiring a skill. While providing this formative feedback is currently the most valuable function of objective assessment with simulation, inevitably, simulators will be used for summative performance assessment. This testing will then be used for processes such selection and credentialing in the future much like knowledge testing is used now. In order for simulators to be applied to such high-stakes assessment it will require a much more rigorous set of metrics and is still in the experimental phase. When this does come to the fore it is certain the metrics for that

simulator must be shown to meet the same psychometric standards of validation as any other psychometric test

1.6.5 Augmented reality simulation

Many modern simulators utilise augmented reality (AR). This term was coined in 1990 and can be defined as an environment that includes both virtual reality and real-world elements. This has many applications in areas as diverse as navigation systems, architectural displays and theme park attractions where the experience of a real physical attraction is augmented by using virtual graphics. With regard to surgical simulators, this often refers to VR capabilities and real physical models, combining the advantages of haptic feedback from a box trainer and virtual graphics and tracking. Simulators which use AR are often referred to as hybrid simulators. The first hybrid simulator to train laparoscopic skills was the ProMIS™ simulator. This simulator consists of a plastic body form which represents the patient's abdomen. Inside the abdominal cavity is a tray with plastic models of the intra abdominal organs. The surgeon trainee inserts their laparoscopic instruments through the foam layer which represents the anterior abdominal wall to manipulate the plastic intra abdominal organs. At this stage the simulator resembles an advanced box trainer. However, what the trainee looks at on the laptop screen is not just the view from inside the abdomen, but also superimposed virtual reality graphics. These graphics, in combination with verbal direction from the simulator guide the trainee through the procedure in a series of steps, demonstrating what action must be taken at every stage of the procedure, ie showing him her exactly where to cut or dissect tissue. A recent study showed that surgeons preferred an augmented reality simulator to a purely VR

simulator (Botden SMB et al, 2007) Examples of physical objects include mannequins or models of intra abdominal organs. The latter may have the added benefit of allowing the trainee to use real surgical instruments while performing tasks on the simulator, enhancing the fidelity of the learning experience.

1.6.6 Simulators as training tools

Despite the advances in the technological side of VR simulators over the past number of years a lack of knowledge about how they should be used could mean that their full benefit would not be realised. Simulators are not an end in themselves but rather a means of delivering a structured teaching curriculum and the most advanced simulator available will not guarantee knowledge acquisition if not used properly.

Simulators train skills but do not always educate. These two terms are often used interchangeably but education refers to the acquisition of knowledge or information while training usually describes skill acquisition. Therefore a simulator alone cannot teach an individual how to do a procedure. The trainee must have knowledge of the relevant anatomy, the steps of the procedure, common errors and how to avoid them, etc. Training on a simulator should only commence when sufficient knowledge has been demonstrated, and therefore most studies using simulators require that subjects have a didactic teaching session and pass an exam or test before proceeding to the simulator training.

There are also other concepts which are important to the successful delivery of a curriculum incorporating simulation. One of the most important is the idea of proficiency-based progression (Gallagher AG et al, 2005). According to this paradigm, the trainee practices on a simulator until a certain, predefined standard has been achieved. This is because individuals learn at different rates, some may acquire a

skill very quickly whereas others may require more training. Therefore, measures that indicate sufficiency of training such as time spent, or amount of procedures performed may not give a true indication of the trainee standard. (Cass OW, 2005) It is a system whereby a trainee is only deemed competent in a particular procedure if they can demonstrate that they can perform the procedure to a certain predefined standard. They can only progress to the next, more complex stage or task when they have attained proficiency in the current task.

As regards surgical skills training, the term proficiency is more appropriate than competency, as proficiency is more concerned with the acquisition of a particular technical skill whereas competency is a broader term referring to the ability to deal adequately with a subject (Gallagher AG et al, 2003).

Proficiency based progression involves setting a benchmark standard from the mean performance of experts in the field. It is important that this standard is not too high, as if so many trainees will not be able to reach it, but high enough to ensure a good, safe standard of operative skill. Proficiency levels for every simulator should be set before training is commenced in order to deliver training more effectively. A trainee could practice on the simulator until they have achieved proficiency, and this would be easily and objectively measured by the simulator based metrics. Different trainees would acquire proficiency after varying amounts of practice, but once they had reached this level, only then would they be permitted to perform the task or procedure in the real clinical setting. Simulation technology lends itself to proficiency based progression.

There has been much resistance to the introduction of simulation to surgical training, due in part to the fact that many surgeons tend to evaluate simulators on a very

superficial level, i.e., does it look like “real surgery”, rather than how the instruments or tissue behave, how appropriate the metrics are, or most importantly how appropriate is the simulation curriculum. It is certainly true that initial simulators were low fidelity but research has shown that this is not an indicator of how effective a simulator is. The best validated VR simulator in medicine i.e., the MIST VR (Minimally Invasive Surgical Trainer in Virtual Reality, uses abstract tasks such as manipulating a sphere in a 3D box to train laparoscopic skills and certainly looks nothing like “the real thing”. It is understandably important that a simulator is fully validated before used in the high stakes environment of surgical training and assessment and much recent research has focused on this. One of the most important kinds of validity is construct validity, ie do the metrics measured by the simulator distinguish experts from novices, and many different VR based simulators have now been validated (Schiven M, 2003).

1.6.7 Effectiveness of simulation – skills transfer to the operating room

While it is intuitive that performance on a simulator will improve with practice, it was important to demonstrate that practicing on a simulator would lead to an improved performance in real surgical procedures. In 2006, Sutherland et al published a systematic review of surgical simulation compared with no training and with other methods of surgical training. Thirty randomised, controlled trials were included and found that despite the safety and ethical advantages that are associated with avoiding patient-based training, simulator based training was not consistently superior to other forms of training. In order to justify the investment in simulation it is important to verify that skills learnt on a simulator will transfer to the real operating environment,

or the real “OR”, hence the term “VR to OR” The first trial to demonstrate this skills transfer from the virtual to the real environment took place in 2001 at Yale University. (Seymour N et al, 2002) A group of residents were enrolled in the study. Half of the resident group was randomized to train on the MIST VR simulator to a level of proficiency and half underwent traditional training. They were then assessed on their performance of part of a laparoscopic cholecystectomy and the residents who had trained on the simulator performed the procedure 29% faster and made 6 times fewer errors. Similar findings were demonstrated in two further studies,(Grantcharov TS et al, 2004; Ahlberg GA et al, 2007) one of which also demonstrated that the performance improvement in the VR-trained group persisted over ten procedures (Ahlberg GA et al, 2007) This shows that simulator-based training can have a persistent beneficial effect when added to the more traditional curriculum. Although each of these trials demonstrated significant differences between the groups’ intra-operative performances, they were limited by small numbers of subjects and two of them only assessed performance of part of a procedure. There have been many other “VR to OR” studies published since then, demonstrating skills transfer. Seymour recently examined VR to OR evidence in the published literature (Seymour N, 2008). Of the 14 studies reviewed, seven studied VR-related acquisition of laparoscopic skills. Six of these showed a positive effect of VR simulator training. Other studies reviewed included trained procedures including flexible bronchoscopy, endoscopy and uteroscopy. Only one of these studies, by Gerson et al (2003) demonstrated a negative effect associated with VR simulation training. In this study, one group of residents was trained to perform a sigmoidoscopy by a training programme involving 10 clinical procedures and compared to a similar group of residents who received VR based training only. Unsurprisingly the group who had trained on the real procedure

outperformed the VR group. This demonstrates the fact that VR simulation is effective when used as complement to but not substitute for traditional “on the job” training. There have been two other recently published comprehensive reviews in the same area. Gurusamy et al (2009) reviewed the effectiveness of VR training for laparoscopic surgery. Twenty three randomised clinical trials were identified. VR based training improved performance in novice trainees when compared with no training, and improved the performance of trainees with limited laparoscopic experience when compared to traditional training. Sturm et al (www.surgeons.org/asernip-s/) performed a systematic review of the effectiveness of surgical simulation training and identified 12 randomised controlled trials and two non-randomised studies. For laparoscopic cholecystectomy, endoscopy, catheter-based intervention for vascular disease and hernia repair, simulation based training showed a benefit. There was no benefit found in a study which used an endoscopic sinus surgery simulator to train residents.

Given the increasing body of evidence supporting its effectiveness, VR simulation training may well become a mandatory part of surgical training, with trainee surgeons only permitted to enter the operating room after they have achieved a certain standard on the simulator, similarly to pilot training in the aviation industry. VR simulation also has applications in assessment and credentialing of surgeons.

Despite initial scepticism, VR simulation looks set to become an integral part of training for procedural based medicine. Specialties which have VR simulators include interventional radiology and cardiology, urology, gynaecology, otorhinolaryngology, orthopaedics, ophthalmology and general surgery, and in 2004, VR simulation

training was approved by the FDA in the United States for carotid artery stenting training (Cates et al, 2003)

1.7 Teaching surgical skills

Although the physical elements of the training environment such as the fidelity of the simulator are important, there are other elements in the training environment which can have an impact on training. These include trainee factors such as self-motivation and fundamental abilities⁸, training factors. In the past, surgical trainees were expected to watch and learn by imitation, but now most teaching programmes use elements of learning theory when designing curricula. The following is a brief overview of learning theories which have relevance to surgical training.

1.7.1 Theories of learning

Behavioural theory.

Behaviourism as a theory was mostly developed by B. F. Skinner. In essence, three basic assumptions are held to be true. First, learning is manifested by a change in behaviour. Second, the environment shapes behaviour. And third the principles of contiguity (how close in time two events must be for a bond to be performed) and reinforcement (any means of increasing the likelihood that an event will be repeated) are central to explaining the learning process. For behaviourism, learning is the acquisition of new behaviour through conditioning. There are two types of possible conditioning: Classical conditioning, where the behaviour becomes a reflex response to stimulus as in the case of Pavlov's Dogs, and Operant conditioning where there is reinforcement of the behaviour by a reward or a punishment

Cognitive theory

Behavioural theories have been criticised for being too dependent on overt behaviour to explain learning. Two key assumptions underlie the cognitive approach: (1) that the memory system is an active organized processor of information and (2) that prior knowledge plays an important role in learning. Cognitive theories look beyond behaviour to explain brain-based learning. Cognitivists consider how human memory works to promote learning. For example, the physiological processes of sorting and encoding information and events into short term memory and long term memory are important to educators working under the cognitive theory. The major difference between Gestaltists and behaviourists is the locus of control over the learning activity. For Gestaltists it lies with the individual learner; for behaviourists it lies with the environment.

Constructivist theory

Constructivism is a theory of learning founded on the premise that by reflecting on one's own experiences, one constructs one's own understanding of the situation or issue encountered. Individuals generate their own "rules" and "mental models," which are used to make sense of or understand their experiences. What any given person observes depends on what is already stored in that person's memory. Our brains process information and images differently because we interpret what we see and hear based on what we already know and expect. This observation suggests that learning occurs within the student's environment (through his or her senses) and is an active, rather than a passive, process. Individuals construct their own meanings shaped by previous knowledge, experiences, and social and economic influences (Benzc 2004) Certain curriculum models promote teacher-centred instruction that treats students as

empty vessels; these models assume that what the teacher transmits is received and understood as intended. Yet empirically we know that deep surface learning occurs only when new knowledge is linked to previous knowledge through the integration of concepts and facts.

1.7.2 Adult learning

Surgical trainees are adults and therefore the field of adult learning is of particular relevance when designing surgical training programmes. Malcolm Knowles' theory of andragogy (the adult version of pedagogy) is an attempt to develop a theory specifically for adult learning. (Knowles M, 1990) Knowles emphasizes that adults are self-directed and expect to take responsibility for decisions. Adult learning programs must accommodate this fundamental aspect. Andragogy makes the following assumptions about the design of learning: (1) Adults need to know why they need to learn something (2) Adults need to learn experientially, (3) Adults approach learning as problem-solving, and (4) Adults learn best when the topic is of immediate value. In practical terms, andragogy means that instruction for adults needs to focus more on the process and less on the content being taught. Strategies such as case studies, role playing, simulations, and self-evaluation are most useful. Instructors adopt a role of facilitator rather than lecturer or grader. (Bruner, 1966) These are all techniques which are employed for example in the Human Factors courses which train "soft" surgical skills. Andragogy assumes that the point at which an individual achieves a self-concept of essential self-direction is the point at which he psychologically becomes adult. When this occurs the individual develops a deep psychological need to be perceived by others as being self-directing. Thus, when he

finds himself in a situation in which he is not allowed to be self-directing, he experiences a tension between that situation and his self-concept.

1.7.3 Theories of motor learning – Fitts and Posner

Fitts and Posner (1967) suggested that the learning process is sequential and that we move through specific phases as we learn. According to this theory there are three stages to learning a new skill:

- Cognitive phase - Identification and development of the component parts of the skill - involves formation of a mental picture of the skill
- Associative phase - Linking the component parts into a smooth action - involves practicing the skill and using feedback to perfect the skill
- Autonomous phase - Developing the learned skill so that it becomes automatic - involves little or no conscious thought or attention whilst performing the skill - not all performers reach this stage

The learning of physical skills requires the relevant movements to be assembled, component by component, using feedback to shape and polish them into a smooth action. Rehearsal of the skill must be done regularly and correctly. For example, with a surgical skill as simple as tying a knot, in the cognitive stage the learner must understand the mechanics of the skill — how to hold the tie, how to place the throws, and how to move the hands. (MacRae et al. 1997) With practice and feedback, the learner reaches the integrative stage, in which knowledge is translated into appropriate motor behaviour. The learner still has to think about how to move the hands and hold the tie but is able to perform the task more fluidly, with fewer interruptions. In the autonomous stage, practice gradually results in smooth performance. The learner no longer needs to think about how to execute this particular task and can concentrate on

other aspects of the procedure.

This model has obvious implications for surgical training. The earlier stages of teaching technical skills should take place in the skills lab; practice is the rule until automation in basic skills is achieved. This mastery of basic skills allows trainees to focus on more complex issues, both technical and nontechnical, in the operating room. To return to the example of knot tying, the learner who still has to think about how to tie a square knot is much less likely to pick up on other teaching that transpires in the operating room than is the learner who has mastered this simple skill. This theory reinforces the idea that only the “pre-trained novice” should be permitted to enter the operating room.

1.7.4 Model of skills acquisition – Dreyfus and Dreyfus model

The Dreyfus Model of Skill Acquisition postulates that when individuals acquire a skill through external instruction, they normally pass through five stages. This model, first proposed by Stuart Dreyfus and Hubert Dreyfus in 1980 proposes that the five stages of skill acquisition are: Novice, Advanced beginner, Competent, Proficient and Expert. A beginner acquiring a skills starts at the bottom and progresses through the 5stages. Recently some researchers have added “master” as a 6th stage. The model was first developed based on work with airline pilots, and has since been notably applied to nursing education (Benner 2004) In the novice stage a person follows rules that are context free and feels no responsibility for anything other than following the rules. Competence develops when the number of rules becomes excessive so organizing principles need to be developed and information sorted by relevance. Competence is

characterized by active decision making. Proficiency is shown in individuals who use intuition in decision making and develop their own rules to formulate plans.

A summary of the model (Eraut, 1994) is below

1 Novice

rigid adherence to rules
no discretionary judgment

2 Advanced beginner

situational perception still limited
all aspects of work are treated separately and given equal importance

3 Competent

coping with crowdedness (multiple activity, information)
now partially sees action as part of longer term goals
conscious, deliberate planning

4 Proficient

holistic view of situation, rather than in terms of aspects
sees what is most important in a situation
uses maxims for guidance, meaning of maxims may vary according to situation

5 Expert

no longer reliance on rules, guidelines, maxims
intuitive grasp of situation, based on tacit knowledge

The progression is thus viewed as a gradual transition from rigid adherence to rules to an intuitive mode of reasoning that relies heavily on deep tacit understanding. The model has relevance to surgical training and can be used to describe the surgical trainee's progression through various stages of competence.

1.7.5 Designing a skills training curriculum

Teaching the physical techniques of surgery or other procedural-based specialities is enhanced by knowledge of learning theories. Some relevant factors are outlined below

A proposed template for developing a curriculum should include the following sequence: (1) didactic teaching of relevant knowledge (ie, anatomy, pathology, physiology); (2) instruction on the steps of the task or procedure; (3) defining and illustrating common errors; (4) test of all previous didactic information to insure the student understands all the cognitive skills before going to the technical skills training and in particular to be able to determine when they make an error; (5) technical skills training on the simulator; (6) provide immediate (proximate) feedback when an error occurs; (7) provide summative (terminal) feedback at the completion of a trial; (8) iterate the skills training (repeated trials) while providing evidence at the end of each trial of progress (graphing the “learning curve”), with reference to a proficiency performance goal that the trainee is expected to attain. While the above is a proposed template, it includes in a stepwise fashion all components published in the literature that would comprise a comprehensive and validated training curriculum. Details of how to create and implement such a curriculum will be discussed below.

1.7.5.1 Task deconstruction

Research supports the approach of teaching complex surgical skills by dividing them into simpler component tasks using cognitive and motor task analysis (Velmahos et al, 2004; Kaufman et al, 1993) This technique also underlies the Fitts-Posner 3 stage learning theory. This view requires a systematic approach to determining steps and

stages of an operation and identifying the preoperative, intraoperative, and postoperative decisions associated with the procedure. The most basic elements of the task can then be individually trained, or shaped (Kazdin AE, 1966) Task deconstruction also facilitates objective assessment..

1.7.5.2 Practice.

The effect of practice on performance is larger than earlier believed. Studies by Ericsson and colleagues (Ericsson et al, 1993) support the critical need for extended, deliberate practice to reach expert performance. Current residency formats move residents through multiple clinical rotations throughout a year. During these rotations, the types of patients encountered depend on the faculty associated with that service. This approach provides experience but does not provide for focused and deliberate practice. Nor does such a rotation system uniformly sequence instruction for efficient skill building.

1.7.5.3 Feedback

Feedback can be defined as the provision or return of performance-related information to the performer. It is recognised as an important part of the learning process in medical education (Ende J, 1983), and can be intrinsic, where it is relayed directly by the sensory system of the trainee, ie the trainee can see if they have made an error, or extrinsic where it is provided by an external source. The medical education literature further describes it as informed appraisal by a teacher, whose aim is to reinforce strengths, highlight the difference between intended and actual results of actions and

provide insight into actions and consequences. Although the benefits of performance feedback are not debated, the optimal type of feedback is not known. Questions have been asked regarding the best frequency (Winstein CJ et al, 1990; Sewell C et al, 2008), and type of feedback, ie simple knowledge of results (O. Connor A et al 2008) vs more intense and escalating feedback (Sickle K et al 2007).

A successful, proficiency-based training programme should incorporate proximate feedback for the trainee (Gallagher AG et al, 2005). One of the most vital differences between a basic box trainer and a more advanced simulator is the ability of the latter to provide performance metrics. These can be crude, such as performance time or more detailed such as instrument smoothness and path length. While these can be used to allow a trainer to objectively assess performance or progress, perhaps more importantly they can be used to give the subject feedback on their performance.

Not utilising this facility therefore means that full advantage is not being taken of the simulator. If the subject is not aware they are committing an error, they may keep doing it, which means that the simulator is reinforcing an undesirable behaviour. Even a high-fidelity simulator may not be an effective training tool in the absence of feedback (Mahmood T et al, 2004)

The advantages of feedback are well recognised in both cognitive and behavioural learning models. Several important studies which use virtual reality simulation to train trainees in a proficiency-based progression programme have used feedback (Seymour N et al, 2002; Ahlberg GA et al, 2007). In these cases trainees were given proximate feedback on their performance after every cycle of tasks, which was repeated until they achieved proficiency. Having attained proficiency in the virtual environment, they then demonstrated a superior performance in the real, clinical environment.

1.7.5.4 Assessment

Closely linked to the theme of feedback is assessment as in the absence of assessment it will be difficult to provide meaningful feedback. Assessment of surgical performance will be discussed in more detail in section 1.8.

1.7.5.5 Curriculum

Based on an extensive review of research on learning and skill acquisition, Ericsson and colleagues (Ericsson et al, 1993) define a number of conditions required for optimal learning and improvement of performance. First, the curriculum must be directed toward highly motivated learners willing to attend to the skill being learned and to exert significant effort to improve performance. Second, the design of the curriculum must take into account the pre-existing knowledge of the learners so that the task can be understood correctly after a brief period of instruction. In other words, learners must be given basic or background material to bring them to a uniform level of understanding before being presented with the skill to be learned. As will be demonstrated in the following experiments, a common way to achieve some uniformity of background knowledge is to incorporate a standardised didactic teaching session into any curriculum. All our didactic teaching was directed at the level of the trainees. Had medical students been involved for example, the didactic would have assumed less background knowledge. Cognitive proficiency can be verified by requiring trainees to pass a cognitive test or quiz before progressing to the practical aspect of the training curriculum. Third, the curriculum must ensure that learners receive immediate informative feedback and knowledge of results of their

performance. Lack of feedback is a common complaint of residents; in the absence of adequate feedback, efficient learning is impossible, and minimal improvement is seen, even in highly motivated learners. Mere repetition of a skill will not automatically lead to a trainee's reaching a high level of skill mastery. Last, the learners should perform the same or similar skills repeatedly. When these conditions are met, practice improves accuracy and efficiency of performance on cognitive, perceptual, and motor tasks. Deliberate practice is a highly structured activity. Focused analysis of problems experienced by the learners when executing the skill provides cues for ways to improve.

1.8 Assessment of surgical performance.

There are many domains of performance which are relevant to surgery. Here we briefly discuss assessment of cognitive and personal factors before focussing on technical skills assessment as it has more relevance to this thesis.

1.8.1 Academic assessment.

Cognitive knowledge is vital for good surgical practice. A thorough understanding of the basic sciences including anatomy, physiology and pathology as well as details of individual operations is extremely important. Cognitive knowledge is one of the domains which is easiest to assess and been assessed in a structured and standardised fashion for many years. It begins at medical school where students must succeed in regular examinations in order to make progress and eventually graduate from medical school. Assessment continues throughout surgical training with examinations administered by surgical training bodies such as the Royal College of Surgeons, such

as membership and fellowship examinations. While these examinations consist of a combinations of written questions, multiple choice style questions and oral exam and provide a thorough assessment of cognitive knowledge, practical aspects of performance are not assessed.

1.8.2 Human factors assessment

Personality, or human factors are also relevant to surgical practice but are rarely assessed. Some assessment takes place at interview when surgical trainees are applying for new posts, but these are infrequent and not designed for the purposes of assessment. The use of 360 degree assessment as described by Whitehouse et al (2007) may provide a means of assessment of non-technical skills, although its validity as a testing tool is to be determined. Other attempts have been made to assess skill domains relevant to surgery such as teamwork behaviours. Mishra et al have developed a rating scale which appears to assess complex team behaviours in the operating room in a reliable and valid manner.

One group used a stress-coping questionnaire to demonstrate that poor-stress coping ability was associated with worse laparoscopic performance in surgical novices (Hassan et al, 2006) . However, Maschuw et al (2008) measured self belief in a small sample of trainees and found that in novices it inversely correlated with laparoscopic surgical performance and that in more experienced trainees there was no correlation. The assessment of non-technical skills in surgery is complex and not yet fully understood but is increasingly recognised as an area which could be beneficial to understand better.

1.8.3 Technical assessment

Until recently, technical surgical performance was never formally assessed. Trainees could complete their training period by passing academic exams or training for a required number of years without any formal assessment of their operative ability. Logbooks are used to examine trainee volume and activity but do not provide a reliable indicator of trainee technical skill. The measure of a surgeon's operative skills can be defined in terms of patient outcomes. Development of more formal surgical training programmes over the past number of years and the increasing use of simulators have made assessment of technical surgical performance both topical and feasible.

1.8.3.1 Assessment of real operative performance -outcome

Success of a surgical procedure can be assessed by measuring certain standard patient-related outcomes. This information is often used to compare the performance of different hospitals and even different surgeons. Typical general measures which are used are indicators of patient's recovery time, which can be a proxy measure of the skill with which the procedure was performed. Standard measurements used are length of stay, post operative length of stay, time to return to work, post-operative pain scores (which may themselves be measured in a variety of ways such as subjective reporting of pain using a visual analogue scale or the duration of requirement for opioid analgesia post-operatively) and quality of life measures at different time-points following the surgery. Such measures are often used for example to compare different types of surgery where the outcome is known to be the same, such as the shorter lengths of stay reported in laparoscopic vs open colorectal surgery (COST study group).

Success of a particular procedure can also be measured and these are more specific to the actual operation carried out. For example most operations for cancer should have recorded oncological clearance, or pathological quality of resected specimens, recurrence rates etc. Other commonly used measures include peri-operative and post-operative mortality, and also morbidity which again may be quite specific to the type of operation carried out. Morbidity can be unhelpful as a comparative outcome measure if it occurs only rarely. For example, comparing the performance of two endoscopists over a 6-month period based on their rate of perforation may not yield meaningful results as the frequency of post colonoscopy perforation is extremely low. Most operative outcomes are superior in higher volume centres and by higher volume surgeons. Such information can be used for internal audit and for individual surgeon feedback. There has been a recent trend for the publication of such data, supposedly to allow patients to make informed choices about where they will have surgery carried out but often driven by other factors such as insurance companies (Lee et al, 2004). This can be misleading as there are a huge number of factors which may influence operative outcome (Lee et al 2004, RCSI eng 2008) such as pre-operative patient health, indicators for surgery etc and there are many instances of inaccurate reporting (Werner et al, 2005).

While there may be a benefit for assessment of hospital/surgeon performance such measures are not suitable for assessment of individual surgeon competence in an examination situation due to patient variability, the need to perform large numbers of operations, the time taken for results to become apparent, and ethical reasons. Such outcome measures, although crude in some ways also give an indication of non-technical abilities of the surgeon as successful operative outcomes do not rely merely on technical skills but on many other factors such as personality and judgement.

However it is almost impossible to know what contribution to the overall outcome relies on each skill domain. In addition, an outcomes-based assessment approach is clearly inappropriate for trainees who are may not be performing complete procedures.

1.8.3.2 Assessment of real operative performance - directly observed assessment

This assessment method involves the direct observation of a surgeon's performance by an assessor. It can be extremely subjective and rater-dependent, and also depends on unavoidable factors such as the patient's health and complexity of the case. To offset some of these difficulties, most directly-observed performance assessments are carried out using a structured assessment tool which is usually specific to the task being performed. The Intercollegiate Surgical Curriculum Programme in the UK has produced procedural based assessment forms (PBAs) for a wide range of trainee operations such as varicose vein repair and appendicectomy. A recent systematic review by Kogan et al for example identified 55 unique assessment tools which were used to grade the clinical performance of students and trainees. An example of a directly observed assessment tool for colonoscopy for example can be found in the index to chapter 7. One method of making the assessment less biased is to observe and assess video-recording of surgical performance with the identity of the operator hidden, This style of assessment is particularly suited to laparoscopic procedures for example where the operating image is already displayed on a monitor and can be recorded with little extra expense. Another way of improving the assessment process is by using a combination of a checklist assessment form and an overall "gestalt" assessment which provides a way to assess elements of performance which are otherwise difficult to measure, such as the system developed by Reznick et al. In

addition, directly observed performance (live or video recorded) can be carried out during the real or simulated patient encounter. Directly observed procedural assessment forms are used for example by the Joint Advisory Group for Gastrointestinal Endoscopy (thejag.org) to credential endoscopists to carry out screening colonoscopies

1.8.3.3 Bench model assessment

As described above, directly observed assessment tools can be easily applied to simulated, bench-model style tasks. This is the basis for the OSATs assessment as described by Reznick et al in 1997 and allows for a thorough trainee assessment in a safe environment with no patient risks. Assessments consist of task-specific checklists which are developed for every station and usually involve task deconstruction analysis. In addition to the checklists, global scoring scales are used for every station. Direct outcomes can also be examined by examining the models after the procedure has been carried out. This is the basis for the error assessments used in Chapter 3. For example, a trainee learning to repair flexor tendons in the skills lab can have the repair assessed by measuring the tensile strength of the repaired tendon in a tensiometer. This is an outcome assessment which cannot be assessed in real life! Similarly, outcomes such as suture strength and anastomotic integrity can be tested easily in the skills lab. In chapter 3 we used a checklist to carry out an error assessment of the anatomy tray following every procedure.

1.8.3.4 Hand motion analysis

This is an objective technique for performance assessment developed at Imperial College London which has the advantage of being objective, and suitable for the

simulated and with some restrictions, real environment. The ICSAD device measures surgeon hand movements during certain tasks using motion detector analysers which are attached to the dorsal surface of the surgeon's hand. The metrics it produces include time taken, and number of movements and these have been shown to be effective discriminators of different skill levels (Datta et al, 2001) and also to correlate with other performance measures (Datta et al, 2002) for both open and laparoscopic tasks. Such measurements would appear to indicate a greater degree of efficiency, control and dexterity by the surgeon.

1.8.3.5 Simulator assessment

One of the advantages of simulators is their assessment capabilities. As discussed in section 1.6, all but the most basic simulators such as box trainers record performance data. This data is objectively measured and can be stored and easily accessed to provide an indicator of surgical ability. In comparison to the ICSAD device, many simulators track the instruments, not the surgeon's hands. Usually the distal tips of any instruments inside the simulator are tracked in 3-D space. Typical metrics include instrument path length for left and right hands, instrument smoothness, instrument angular path length and procedural time. Other metrics which are specific to the simulator or procedure can be assessed, and typically these metrics have content or face validity. For example, the Lap Sim simulator (described in detail in Chapter 2) measures general instrument movements. It also measures errors specific to the performance of the task in question (laparoscopic cholecystectomy) such as mean burn time, energy burnt in air, liver blood loss, etc. Some metrics are generally regarded as less important in performance. Procedural time for example can be a differentiator of skill when all the other scores are equal but should not be regarded as

an important metric. While metrics such as these have been shown to possess construct validity (Pellen MG et al 2009, Neary PC et al 2008) they are less relevant to real clinical practice and outcome measures or error assessments may be more relevant. Where the trainee is practising a basic task and ease and fluidity of instrument movement are valuable metrics, but error scores may be more relevant, and this is discussed further in Chapter 3. There is a danger that metrics such as path length and smoothness are measured because they can be measured. It is also important that trainees treat tasks using these measurements in a realistic manner. For example, a trainee performing a simulated task which involves touching the instrument against various targets will achieve better scores for path length and smoothness if he or she does not try to hit all the targets. In such cases, if the amount of targets achieved is not assessed also, a trainee could artificially improve their instrument movement scores by choosing to hit fewer targets. However if the trainee treats the task as a real surgical case, the metrics may give valuable feedback and may help to discourage unnecessary and extraneous movements.

To summarise, when taken in context, simulator-based metrics provide an objective and standardised way of assessing and tracking trainee performance.

Chapter 2

Materials and Methods Overview

Chapter 2 Materials and Methods overview

2.1 Overall aims of the studies

The overall aim of the thesis was to identify ways to improve simulation-based surgical training. We have specifically not set out to demonstrate the benefit of simulation training per se as this has already been established in the literature (Seymour 2008, Gurumay et al 2009). We aimed to investigate ways in which the benefit of simulation-based surgical training can be maximised. In order to do this we have selected 3 image-guided procedures on 3 separate high –fidelity virtual reality simulators – hand-assisted laparoscopic colectomy (HALC), renal artery angioplasty and stenting and colonoscopy. In addition, we have used basic tasks on other simulators and training tasks on the Nintendo Wii to further investigate training issues. A full description of the simulators used is detailed in section 2.4 Although we have used separate studies and procedures, the experiments in the thesis follow a clear pattern. All studies involve participants who are surgical trainees of varying levels, but novice in the particular procedure with the exception of the trainees in the Wii study who were medical students. All studies compared control and test groups. The first study (Chapter 3) investigates the value of structured feedback vs no feedback for trainees performing a simulated HALC procedure. Learning curve patterns and the relevance of different performance metrics are also discussed. A small substudy also examines skills transfer to different tasks, ie does practising one procedure improve performance in a separate procedure. The next study (Chapter 4) develops this theme further by investigating the benefit of feedback provided by different observers, comparing expert to non-expert to no feedback. The aim of this section was to assess

the necessity of providing expert clinical faculty at every training session. This study uses the renal artery stenting procedure. The next study (Chapter 5) investigates the effect if any that practising on a non-surgical simulator, ie the Nintendo Wii may have on surgical skills as has been suggested in the literature. Chapter 6 examines retention of surgical skills using the HALC and renal artery procedures and tests retention of surgical skills at different time points. In Chapter 7, the impact of a structured, simulation based curriculum for colonoscopy is examined, incorporating some of the principles already discussed such as cognitive teaching, provision of feedback, proficiency-based progression and skill retention. Chapter 8 investigates the topical relationship between innate abilities and their relationship to surgical skills. It has been suggested that surgical training can be optimised when those trainees most suited to the demands of surgery are selected preferentially and the measurement of innate abilities has been suggested as part of the selection process.

The aim of the research was to identify factors which may have relevance to the delivery of surgical training in the current era of simulation based training and to examine these factors with a view to improving the design and delivery of surgical training.

2.2 Subjects

All subjects in the studies participated voluntarily. Subjects who participated included medical students, surgical trainees (at senior house officer (SHO) and registrar level) and senior surgical trainees (registered on specialist surgical training schemes).

However all subjects were novice in the procedures that they performed during the course of the studies. Given the unstandardised nature of current surgical training it is difficult to ensure a similar ability level at baseline of the trainees. It is also unrealistic

and does not reflect the current clinical situation. All the studies had specific inclusion and exclusion criteria in order to minimise any existing differences in ability level at baseline and there were no baseline significant differences in fundamental abilities between any of the trainee groups. The training that every trainee received in the various studies was standardised.

2.3 Methodological issues

2.3.1 Ethical approval

All of the simulator-based studies received ethical approval from the RCSI Research and Ethics Committee prior to commencement. Ethical issues that were raised during the course of the application process revolved around trainee confidentiality and data protection. All trainees were recruited voluntarily, through a combination of web-based notices, bulk emails and letters and posters. All trainees were supplied with detailed information sheets before they agreed to become involved. Although in all cases trainee identities could not be hidden during practical data gathering/ training and assessment sessions, all trainee data which was recorded was done so anonymously using random codes. Where trainees were video-taped (Chapter 4) this was done so anonymously. Data was recorded solely for the purposes of the studies and not shared with any third parties. Specifically no trainee data was made available to individual trainers or hospitals. For Chapter 7, individual hospital ethical approval was obtained before commencement, with the same confidentiality procedures for patients. Consent was specifically sought from patients and it was emphasised that their participation on the study would not impact on their hospital care. In most cases for the colonoscopy study, the procedure was carried out by a trainee who was

training in that particular hospital and could realistically have been carrying out that particular case regardless of their participation in the study.

2.3.2 Training

Training in all the studies was carried out at the National Surgical Training Centre (except for the comparison group in Chapter 3). Training was carried out in a standardised fashion for all participants and consisted of massed training on most occasions. Periodicity of training was similar for the trainees in different groups.

2.3.3 Assessment and endpoints

In the majority of cases assessment was carried out using simulator metrics, providing an objective and valid means of assessing performance. Individual assessment metrics are discussed in section 2 xxx and the individual chapters, and are based around instrument tracking technology which is incorporated into each simulator. In addition these metrics were supplemented by additional assessments, such as the tray error assessments in chapter 3 and the video assessments in chapter 4, both of which were carried out in a blinded fashion using standardised assessment tools. With regard to the colonoscopy study in chapter 7, the endpoint was clinical performance of a real colonoscopy which was judged using a directly observed procedural assessment, and although the assessors were not blinded to the identity of the trainee, they were blinded to the trainee's training status, ie control or study group. In chapter 3 and chapter 7, subjective assessments were carried out by two assessors with inter-rater reliability recorded.

2.3.4 Statistical analysis

SPSS Version 15.0 was used in all cases for analysis. For most studies, data were entered into Microsoft Excel spreadsheets and then transferred to SPSS data files. Most data recorded were numerical, of an interval nature and would have resembled a normal distribution. However in some cases data sets were small and in these cases non-parametric tests were used. Where there was any doubt about the appropriateness of which test to use, Shapiro Wilks tests were used to check for normality of the data. For assessment of differences between two groups, paired T-tests or Mann-Whitney-U tests were used as appropriate. For comparisons between 3 groups, ANOVA or Kruskal-Wallis tests were used as appropriate. In these situations, post-hoc testing was carried out using Bonferroni or paired Mann-Whitney-U tests respectively. Learning curve data were analysed using ANOVA for repeated measures and Friedman tests. Learning curves were compared using 2-factor ANOVA. Most correlations were carried out using Spearman correlation coefficients as different assessments were under comparison. Inter-reliability was measured using formula $\frac{\text{number of agreements}}{\text{number of possible agreements}}$ and a minimum value was set at 0.8. In all cases, a p value < 0.5 was regarded as significant.

2.4 Description of the simulators used in the studies

2.4.1 GI Mentor (Simbionix, Cleveland, USA.)

The GI Mentor consists of a plastic bodyform with a trunk and alternating headpiece/tailpiece which is changed depending on whether upper or lower gastrointestinal endoscopy is performed . The endoscope used is a customized Pentax ECS-3840F endoscope. This has the normal controls found on an endoscope, including 2 steering wheels, buttons for air/ water and suction functions, a button for taking pictures and an instrument channel with real instruments which can be used to

perform therapeutic procedures depending on the clinical scenario. This is inserted either into the upper or lower end of the bodyform depending on the procedure. Once inside the plastic bodyform, the colonoscope is traced using sensors so that any movements which are made by the trainee are picked up. There is also a computer monitor and keyboard attached to the simulator. The monitor displays the endoscopic image during the case just as would take place in the real clinical setting.

Figure 2.1 – the GI Mentor Simulator



The simulator allows trainees to perform basic psychomotor tasks which involve navigating the colonoscope along a virtual tunnel without touching the walls, and performing various tasks such as placing a target object into a basket. However the principal aspect of the simulator is a range of upper and lower GI tract endoscopies, including esophagogastroduodenoscopy, endoscopic retrograde cholangiopancreatography, sigmoidoscopy and colonoscopy.

For colonoscopy for example there are 20 separate cases with clinical histories and investigation results which are displayed before each case.

When a particular case is selected, clinical information is given and then the trainee inserts the colonoscope. What they encounter depends on which case has been selected, for example case 5 contains a tumour and an arteriovenous malformation. The trainee manipulates the colonoscope to navigate their way around the colon until they reach the cecum. They may have to perform therapeutic manoeuvres as they do this depending on the case selected. The image of the inside of the colon which is displayed on the monitor is interactive and for example deforms on contact with the colonoscope. In addition the trainee can view a 3-d map during the procedure if they have difficulty in progressing. This gives a real time picture of the colonoscope. When they have reached the cecum, the colonoscope is then withdrawn, as in a real case. However, when we used the simulator to train subjects, we required them only to reach the cecum, and did not assess the withdrawal part of the case. This is discussed further in chapter 7.

The simulator records a range of parameters for each exercise, which can be used to assess performance objectively. These parameters include time to reach the cecum, percentage of mucosal surface examined, amount of time the patient is in pain, number of times the lumen is not visualised and number of occasions of excessive pressure. All of these parameters are relevant to real life colonoscopy. Patient pain, for example is calculated by taking into consideration several factors, such as: shape of the colon, how much scope is currently inside the patient, sudden movements of the scope, etc. These are the parameters that may cause patient pain in real-life.

In addition to the view from the endoscope, the time taken for the case (beginning from insertion of the endoscope) and an indicator of patient pain levels is displayed on the monitor so the trainee is aware of these parameters as they perform the case. The expression on the patient face indicates pain levels, when the pain reaches a certain

severity the “patient” shouts, again mimicking the real life situation. Trainees using the simulator can login every time they have a practice session, meaning that all their performance results are saved and can be accessed to assess trainee progress.

We chose a selection of seven metrics to use in the study, based on clinical relevance and realism. These included time to cecum, percentage of mucosa examined, percentage efficiency of screening, percentage of time spent with a clear view of the lumen, number of times the luminal view was lost and number of occurrences of excessive local pressure. These metrics were used when we set proficiency levels, and were thus used as a benchmark standard that trainees had to reach before their training was complete.

2.4.2 Procedicus VIST™ (Vascular Interventional Surgical Trainer) system (Mentice, Gothenberg, Sweden)

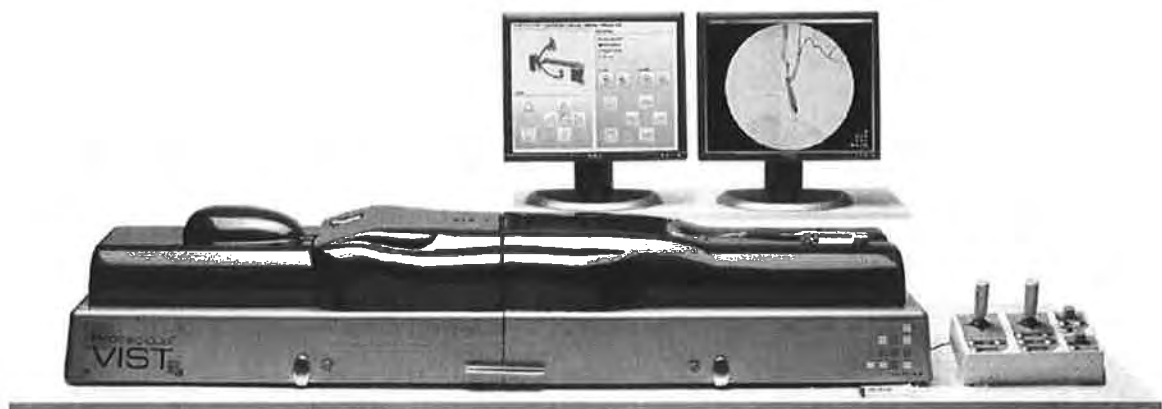


Figure 2.2 Procedicus VIST simulator

The VIST is a high-fidelity, advanced virtual reality simulator which allows the trainee to perform and practice complete endovascular procedures. It consists of a mechanical interface device which is a large, plastic bodyform and represents a supine patient, a high-performance desktop computer, and two display screens (Fig 2.2). The simulator allows trainees to perform complete endovascular procedures. The simulation software allows the selection of dozens of different carotid, renal, iliac, and superficial femoral artery scenarios. The interface device is designed as the virtual patient with simulated right femoral arterial access. Three separate haptic units allow the introduction of real guidewires, guiding catheters, and balloons/ stents into simulated images.

Standard interventional catheters and wires are used with the tips removed. Once inserted, the type of instrument tip required can be selected, ie a 6F diagnostic catheter is inserted with the tip removed, and this can represent a pigtail catheter, Judkins catheter, a Cobra, etc depending on which option is selected.

Separate controls are provided for stent deployment, contrast injection, fluoroscopic C-arm and table movement, cine video runs, roadmapping capabilities, and measurements.

The trainee can select a region and clinical scenario and then insert the various instruments into the plastic bodyform while looking one of the display screens. The display screen represents the fluoroscopic images that would be generated by the X-ray machine if the case were being performed in a real-life setting. They then perform the case. Clinical anatomy is very realistic, even if a particular area is not relevant to the case being performed. For example, the trainee may be attempting to inject contrast into the renal vessels in a renal artery case, but if they inject contrast in the

wrong location, various other vessels will be displayed, such as the superior mesenteric artery. Most cases require the trainee to diagnose and then treat a specified lesion. An assistant can control the C-arm position and other controls, and this is more similar to the real –life setting but it is also possible to perform a procedure unassisted.

Various objective measures of the quality of the procedure such as volume of contrast fluid used, fluoroscopy time, and markers of stent placement accuracy are recorded and tabulated for each participant by the system software.

Trainees can be logged into the simulator and all their performance data saved allowing for continuous assessment. Similarly, specific courses and curricula can be designed.

2.4.3 The ProMIS Laparoscopic Colectomy simulator (Haptica, Dublin, Ireland)

The ProMIS augmented reality simulator trains laparoscopic surgical techniques through a selection of basic tasks and full procedures. It functions as a box trainer, but with virtual reality graphics which are displayed on a laptop screen which represents the laparoscopic monitor. In addition instrument movements are tracked so that performance can be objectively assessed. Precise measurements of time, instrument path length, and smoothness of movement — detected by changes in instrument velocity and direction — are recorded for operating instruments during each procedure.



Figure 2.3 – ProMIS simulator

There is a large range of basic tasks which train the skills required to perform complete procedures. Some of these tasks are purely virtual, such as the instrument holding and positioning task which requires the trainee to touch and hold steady on various targets which are positioned around a virtual abdomen. This task does not offer any haptic feedback but is useful to acclimatise novices to the fulcrum effect and ergonomics of laparoscopic surgery. Other basic tasks involve the use of a real model which is inserted into the bodyform such as for the bead transfer task, in which trainees transfer beads to and from different containers. The simulator provides verbal cues to guide the trainee and onscreen graphical instruction. Other skills which are trained in the basic tasks include fine dissection, clipping and grasping.

To perform a complete hand-assisted colectomy (HALC) procedure, a synthetic anatomy tray (Limbs and Things, Bristol, UK) with anatomically correct models of

the intra-abdominal viscera is inserted into the simulator. A complete procedure can then be performed, divided into 9 steps for teaching purposes.

1. Tilting Bodyform
2. Identification of the inferior mesenteric artery (IMA)
3. Identification of the left ureter
4. Transection of the inferior mesenteric vein (IMV)
5. Mobilisation of sigmoid colon
6. Mobilisation of descending colon
7. Take down of splenic flexure
8. Creation of anastomosis
9. Testing of anastomosis



Figure 2.4 Synthetic anatomy tray

All of these steps are important aspects of the real procedure although the simulator starts from the point at which all incisions/port sites have been created and pneumoperitoneum established. As for the basic tasks, verbal instructions are provided for each step. In addition, when the simulator is in “Teach me” mode, there are on-screen graphical instructions which demonstrate each step and show the trainee

how to perform each aspect of the procedure. The trainee can also select “Test-me” mode, in which the on screen graphic cues are removed. There are also short video clips on the simulator screen which show real-life demonstrations of each step which promotes clinical correlation.

The trainee uses real surgical instruments to perform various aspects of the case. For example, for step 4 they physically transect the plastic vein with a real linear stapler.

Every time a trainee logs onto the system a record of their performance is stored on the database, thus providing an objective record of their progress.

The simulator measures performance using the same instrument metrics as for the basic tasks, ie time, instrument path length (IPL) and instrument smoothness (IS). In addition, the anatomy tray can be assessed for the presence of particular errors after every procedure. For the purposes of the study, we used a predefined set of 14 errors. Thirteen of these errors have been shown to be construct valid from a previous study (Neary et al, 2008), using the same simulator but performing a laparoscopic colectomy. The last error, 14, was added subsequently as in the previous study it had been enacted by some of the subjects. These errors are clearly and explicitly defined (Table 2.1)

Table 2.1 Description of anatomy tray errors

Anatomy tray metric errors	Error definition
Inadequate division of inferior mesenteric vein	<ul style="list-style-type: none"> • divided between its origin at the aorta and its first branch (the left colic artery); • transected and the ends sealed with either a complete staple line or two laparoscopic clips; • divided completely and both free ends separated without any residual tissue remnant connecting the free ends.
Inadequate division of inferior mesenteric artery	<ul style="list-style-type: none"> • transected completely; the free ends should not be in continuity and there should not be any remnant tissue connecting the divided ends; • sealed using a linear stapler, laparoscopic clip applicator or energy delivery device; • [the point of division of the vein is not] in proximity to the point of division of the inferior mesenteric artery and proximal to the first tributary from the colonic mesenteric arcade.
Mesenteric injury	An error will be recorded if a staple line is applied across the mesentery
Inadequate exposure of left ureter	An error will be recorded if there less than 1cm of the left ureter exposed
Inadequate division sigmoid mesentery	An error will be recorded if the sigmoid mesentery is <u>not</u> divided from the point of division of the inferior mesenteric artery to the start of the pelvic brim i.e., incomplete or discontinuous line of division running from a proximal point of origin located at the point of transection of the origin of inferior mesenteric artery to the

	<p>junction of the rectosigmoid mesentery.</p> <p>The end point of the line of transaction should not be greater than 5cm from the medial aspect of the sigmoid colon and not be greater than 10cm from the right aspect of the pelvic brim.</p>
Inadequate mobilisation of left colon	<p>An error will be recorded if the lateral attachments of the colon are not completely divided from an origin point located 10cm from the left aspect pelvic of the pelvic brim (origin of the left lateral fold of the mesorectum) to within 10cm of the proximal lateral attachments of the descending colon at the splenic flexure (reflection of colonic mesentery from the retroperitoneum of the left paracolic gutter termed the line of Toldt). There should not be any residual retroperitoneal attachments to the colon to prevent complete lateral mobilization following this dissection.</p>
Inadequate mobilisation of splenic flexure	<p>An error will be recorded if the mesentery of the splenic flexure is not divided completely permitting the mobilization of the splenic flexure of the colon to be “brought down”. The mesenteric attachments at the splenic flexure should be divided to expose the underlying retroperitoneum. The line of division should incorporate the point of completion of the lateral colonic mobilization and extend for a minimum of 10cm to release the attachments of the distal aspect of the transverse colon. The line of dissection should be complete and continuous to facilitate the mobilization of the splenic flexure.</p>
Inadequate division of mesorectum	<p>An error will be recorded if the mesorectum is <u>not</u> divided from the point of completion of the medial division of the sigmoid mesentery to the mid rectal point. The line of dissection should commence at the junction of the sigmoid mesentery and the origin of the mesorectum. This may be a variable point of origin as it is dependant upon the point of</p>

	<p>completion of the previous medial dissection of the sigmoid mesentery. The line of dissection should extend parallel to the rectum for a minimum of 5cm and leave a >3cm fold of mesorectum intact around the upper rectum. The point of completion of the mesorectal dissection is from the terminal portion of this lateral dissection to within 1cm of the right lateral border of the rectum itself. Similarly, the mesorectum should be divided on the contra lateral side to complete mobilization of the mesorectum.</p>
<p>Inadequate rectal transection</p>	<p>An error will be recorded if the rectum is not transected completely in a line perpendicular to its longitudinal axis. The staple line should be complete and the lumen of the rectum should not be visible. The staple line should not result in a dog ear” deformity. This occurs when the stapler as been applied for a second time however the line of transaction now is aimed proximally towards the sigmoid colon. The resultant “dog ear”” is a potential point of ischaemia and a technical error. The staple line should not include the ink marked colonic lesion. The mesorectum should not be incorporated into the stapling line.</p>
<p>Inadequate anastomotic alignment</p>	<p>An error will be recorded if the orientation of the descending colon and rectum is not preserved in the final anastomosis. The colorectal anastomosis should not be twisted or lie in the form of a spiral. This should be confirmed by maintaining the rectum in its normal orientation and thereafter identifying the orientation of the descending colon and tracing this orientation down to the colorectal anastomosis itself. The correct orientation of the mesentery of the descending colon maintains the mesentery located at the inferomedial aspect of the bowel. A twist or spiral of the mesentery will result in torsion of</p>

	the anastomosis when traced down to the anastomosis itself.
Anastomotic tension	An error will be recorded if the anastomosis is under excessive tension. This may occur when the lateral mobilization is incomplete, the splenic flexure has not been mobilized or there has been a very large specimen length extracted without adequate mobilization. The anastomosis should be capable of being lifted from the retroperitoneum without any tension for a minimum of 10cm height. The anastomosis should be capable of displacement in line with the colon for a length of 10cm in the longitudinal axis.
Anastomosis not centred	An error will be recorded if the colorectal anastomosis is not centred in the middle of the rectal transection staple line. The anastomosis should incorporate the rectal transection staple line itself. The anastomosis should not incorporate the terminal ends of the rectal staple line.
Visceral injury	An error will be recorded if there has been disruption, transection or perforation, as characterized by creation of a defect in the structural integrity of the anatomy tray, of the serosa or adventitia of a hollow viscus (e.g., colon, rectum, stomach), or vessel (e.g., aorta, iliac arteries, ureters).

2.4.4 LapSim (Surgical Science, Gothenberg, Sweden)



Figure 2.5 - LapSim

The Lap Sim surgical simulator is a purely virtual simulator. It consists of 2 generic laparoscopic instruments mounted on a frame which have 360 of rotation and a screen which displays the instruments in a virtual environment. There is no haptic feedback but the movements made by the instruments are viewed on the monitor. A range of basic tasks can be performed, in addition to laparoscopic cholecystectomy and gynaecological procedures

LapSim Cholecystectomy simulates the critical steps in the frequently performed laparoscopic cholecystectomy procedure (Figure 2.6). In the first part, the cystic duct and artery are clipped and cut off. In the second part, the gall bladder is separated and removed from the liver.

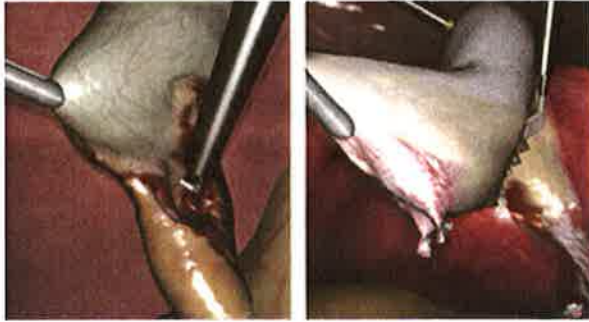


Figure 2.6 Gallbladder task, Lap Sim

2.4.5 Nintendo Wii video game

The Wii is a home video game console released by Nintendo. The hardware includes a remote controller, sensor bar which detects the movements of the controller, console (similar in size to 3 DVD cases stacked together) and Nunchuk, which is like a second remote. A monitor is also required to use the hardware.

A distinguishing feature of the console is its wireless controller, the Wii Remote, which can be used as a handheld pointing device and detects movement in three dimensions (Figure 2.7). The Wii Remote is the primary controller for the console. It uses a combination of built-in accelerometers and infrared detection to sense its position in 3D space when pointed at the Sensor Bar. This design allows users to control the game using physical gestures as well as traditional button presses. The controller connects to the console using Bluetooth . The Wii Remote can connect to expansion devices through a proprietary port at the base of the controller. The device bundled with the Wii retail package is the Nunchuk unit, which features an accelerometer and a traditional analog stick with two trigger buttons (Figure 2.5). In addition, an attachable wrist strap can be used to prevent the player from unintentionally dropping or throwing the Wii Remote.

So for example, when playing a golf game, traditional video games require the player to press a button to control the power of the swing, etc but the Wii allows the player to perform the physical gesture of swinging the golf club by while holding the Nunchuck or Wii remote. The same design applies to other sport games such as boxing (Figure 2.8). For the study we performed, we trained subjects using a gaming software DVD called “Super Monkey Balls”



Figure 2.7 – Wii Nunchuck and remote

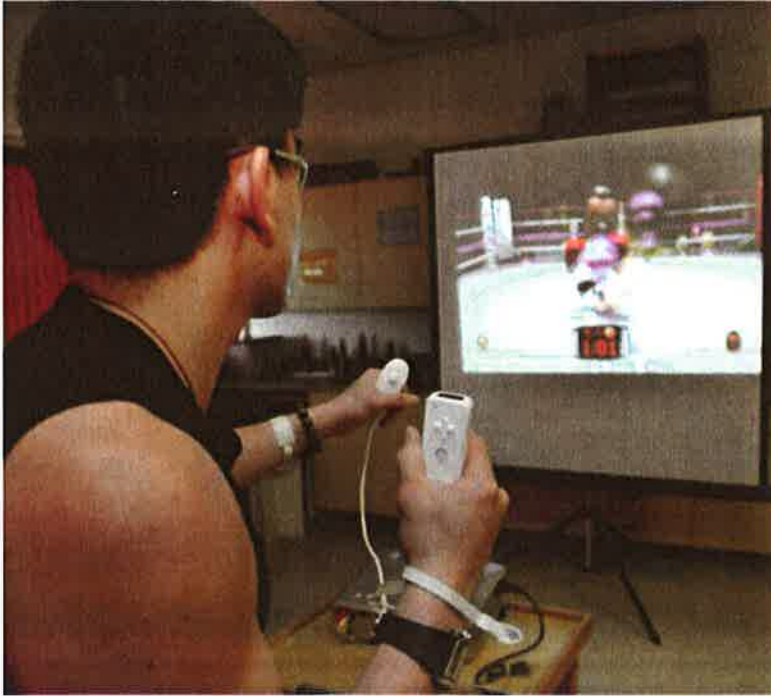


Figure 2.7 – The Wii in use

2.5 Description of fundamental skills testing

Subjects who participated in the three major experiments discussed in the thesis had to complete a series of tests of their fundamental abilities. We used tests designed to measure ability in three domains – visuospatial recognition, perception, and psychomotor ability.

2.5.1 Visuo-spatial

Visuo-spatial ability was measured with the cube-comparison, card rotation and map planning tests taken from the Kit of Factor-Referenced tests (Ekstrom et al, 1976).

These are paper-and-pencil tests which are designed to measure various innate abilities such as reasoning, verbal ability, spatial ability, memory, and other cognitive

processes. The complete kit contains 72 tests that have been demonstrated to be consistent markers in studies of 23 cognitive factors. The kit tests are intended for research use only. Each of the three tests we used consists of 2 sections which have a time limit for completion of 3 minutes each. In addition, each test has an initial section which gives a detailed explanation of the test, and provides some samples for the subject to work through. In order to standardise the pre test instructions, we allowed all subjects to read the supplied instructions independently, and then answered any queries they had before commencing the test.

The Card Rotation Test. (CRT) This requires the subject to compare 2-d figures which may have been rotated into a different position within the plane of the page (same) or rotated through 180°, ie outside the plane of the page (different). In each case, the subject must compare 1 figure with 8 figures which are presented in different positions and indicate if they are the same or different. There are 80 shapes per page, and incorrect answers are negatively marked.

Cube Comparison Test (CCT). This test requires the subject to compare pairs of cubes and indicate if they are the same or different. Each cube has 6 different faces, and as the cubes are presented at different angles, the subject must decide if it is the same cube in a different position, or a different cube. There are 42 pairs of cubes to compare, and this test is also negatively marked.

Map Planning Test (MPT). This test requires the subject to navigate the shortest route possible between 2 points on a grid with certain obstacles to avoid. There are 40 routes in total and the test is not negatively marked.

2.5.2 Psychomotor testing

Psychomotor ability was measured by performance on the ProMIS laparoscopic surgical simulator (Haptica, Dublin). The ‘Instrument Handling-Locating & Coordinating’ module was used. In this module, the bi-manual use of laparoscopic instruments is tested. Two real instruments (right & left hand) are used to touch targets and maintain a steady position in a virtual abdomen. Once inside the bodyform all instruments are tracked in three dimensional space, generating assessment metrics. These metrics comprise time taken, instrument path length (distance travelled by the distal tips of the operating instruments, measured in mm) and instrument smoothness scores (number of accelerations and decelerations, measured in counts), which are objectively recorded and saved by the simulator.

2.5.3 Perception

Perceptual ability was measured with the Pictorial Surface Orientation (PicSOr) “cube and arrow” test (Cowie, 1988). This computer-based test works by generating two objects; a target cube and a rotating probe that connects by its tip with the slanting upper surface of the cube. The aim of the task is to set the rotating arrow so that it appears to be at a 90-degree angle to the surface of the cube. The subject tilts the arrow forwards or backwards until they are satisfied with its position relative to the surface of the cube. This requires the subjects to reconstruct three dimensional objects from their two dimensional representations with minimal depth cues. Before the test starts subjects are allowed to perform several trials of the test and are given their results after every attempt so they know how close their estimate was. When the test commences, subjects are not given any performance-related feedback. They are required to complete 35 trials in total and the test is not timed. When they have

completed the test, their score is calculated by performing a correlation between their estimate and the actual result in every case. This generates a performance result between 0 and 1. The test has previously been validated as a predictor of endoscopic simulator performance.

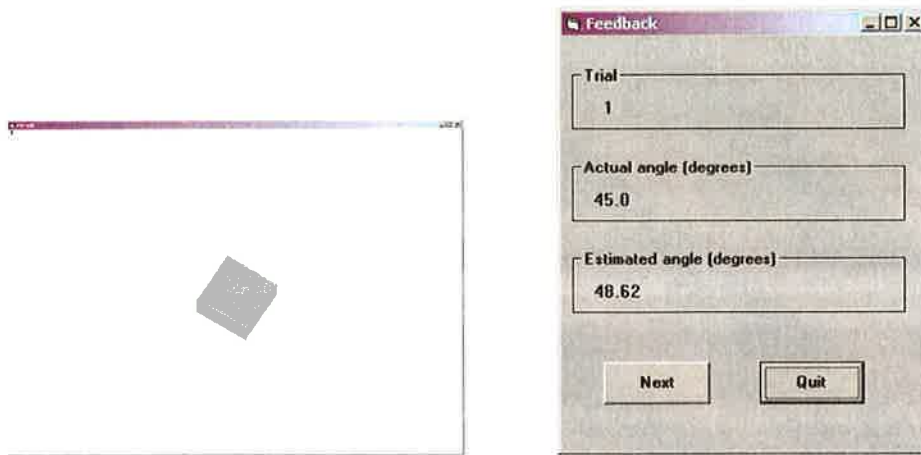


Figure 2.7. Screen shots from the PicSOR perceptual test

Chapter 3

**Improving surgical training: feedback reduces
error scores in simulated hand-assisted
laparoscopic colectomy**

Chapter 3 Improving surgical training: feedback reduces error scores in simulated hand-assisted laparoscopic colectomy

3.1 Background

The last few years have seen huge changes in surgical training (O'Shea J, 2008). Some of the factors which are driving this are increasing complexity and specialisation in surgery, the reduction in the number of training hours available to junior doctors as a result of various working hour restrictions (www.dohs.ie, ACGME), shortages in operating room time, and an increased awareness of the morbidity that can result from operative errors (Kohn LT et al, 1999). As a result surgical training must become more standardised, efficient and effective. Removing certain elements of surgical training from the clinical environment may resolve some of these issues (Mac Rae H et al, 2006).

One important development is the introduction of virtual reality (VR) simulation to surgical training. Although well established in other fields such as aviation and the military, simulation was first proposed as an adjunct to surgical training in 1993 (Satava RM, 1993), and has only recently been given serious consideration as a training tool in surgery. The advantages of the system are obvious; the surgeons can acquire and train basic skills in the safe environment of the simulation lab, supplanting the early, steep part of the learning curve and protecting the patient from unnecessary risk as they overcome the human factor and ergonomic difficulties related to challenging surgical procedures. This also saves valuable and scarce theatre time. The rapid growth in the area of minimally invasive surgery (MIS) over the last two decades has also served to drive simulation as a valuable training tool. Reports of complications early in the

learning curve for laparoscopic procedures (The Southern Surgeons' Club 1991, Dunn et al 1994) reinforced the fact that MIS requires a different set of skills from that required for open surgery. The image guided nature of laparoscopic surgery makes it an ideal system to represent with a simulator (Akay M, 2001), and most VR simulators today train laparoscopic skills. Importantly, skills acquired on a simulator have been shown to transfer to the real environment of the operating room (Seymour N et al, 2002; Grantcharov TS et al, 2004; Ahlberg GS et al, 2007; Chaer RA et al, 2006)

Although the face validity and fidelity of the simulator are important, there are other elements in the training environment which can have an impact on training. These include trainee self-motivation, fundamental abilities, the chance for each trainee to progress at their own pace, and the provision of immediate, or proximate feedback. Feedback can be defined as the provision or return of performance-related information to the performer. It is recognised as an important part of the learning process in medical education (Ende J, 1983), and can be intrinsic, where it is relayed directly by the sensory system of the trainee, ie the trainee can see if they have made an error, or extrinsic where it is provided by an external source. The medical education literature further describes it as informed appraisal by a teacher, whose aim is to reinforce strengths, highlight the difference between intended and actual results of actions and provide insight into actions and consequences. Although the benefits of performance feedback are not debated, the optimal type of feedback is not known. Questions have been asked regarding the best frequency (Winstein CJ et al, 1990; Sewell C et al, 2008), and type of feedback, ie simple knowledge of results (O. Connor A et al, 2008) vs more intense and escalating feedback (Sickle K et al 2007).

A successful, proficiency-based training programme should incorporate proximate feedback for the trainee (Gallagher AG et al, 2005). One of the most vital differences

between a basic box trainer and a more advanced simulator is the ability of the latter to provide performance metrics. These can be crude, such as performance time or more detailed such as instrument smoothness and path length. While these can be used to allow a trainer to objectively assess performance or progress, perhaps more importantly they can be used to give the subject feedback on their performance. Not utilising this facility therefore means that full advantage is not being taken of the simulator. If the subject is not aware they are committing an error, they may keep doing it, which means that the simulator is reinforcing an undesirable behaviour. Even a high-fidelity simulator may not be an effective training tool in the absence of feedback (Mahmood T et al, 2004)

The advantages of feedback are well recognised in both cognitive and behavioural learning models. Several important studies which use virtual reality simulation to train trainees in a proficiency-based progression programme have used feedback (Seymour N et al, 2002; Ahlberg GA et al, 2007). In these cases trainees were given proximate feedback on their performance after every cycle of tasks, which was repeated until they achieved proficiency. Having attained proficiency in the virtual environment, they then demonstrated a superior performance in the real, clinical environment.

The aim of this study was to assess the impact of structured, proximate feedback on surgical trainees' performance of a simulated Hand-assisted Laparoscopic Colectomy (HALC) procedure on the ProMIS simulator, focussing on error scores. Secondary aims were to examine the learning curve for the procedure, assess the correlation between the two main assessment parameters on the simulator, ie instrument efficiency scores and error scores, and in a subset of the subjects, assess for skills transfer between different image-guided procedures, in order to determine if the

intense HALC training provided would improve subject's performance of other procedures.

3.2 Materials and Methods

3.2.1 Ethical approval

Institutional ethical approval was obtained before the commencement of the study

3.2.2 Subjects

28 subjects were recruited to the study. All subjects were surgical trainees of post graduate (clinical) year (PGY) 3-5. No subjects had previously performed any HALC, laparoscopic colectomy, or advanced laparoscopic procedures.

3.2.3 Simulator

The PromIS augmented reality simulator (Haptica, Dublin, Ireland) used in this study is an augmented reality simulator, and is described in chapter 2.

3.2.4 Study Procedures

Training and testing for each subject took place within one day at a centralised location. All subjects signed consent forms before participating and supplied basic demographic data.

3.2.4.1 Teaching session

All subjects were given training on how to complete a simulated HALC procedure using a brief, standardised curriculum incorporating:

- Simulated HALC procedure steps via video/live performance of a simulated case with commentary by an experienced surgeon
- Explanation of performance metrics (i.e., instrument path length, instrument smoothness)
- Explanation of intra-operative errors with examples of errors committed on anatomy trays
- Use of laparoscopic instruments (ETS 45mm linear stapler, and ACE Harmonic Scalpel)

After the formal teaching, subjects took a short true/false test to ensure the course content was understood.

All subjects then performed some brief tasks designed to acclimatise the subjects to the simulator and allow them to use the laparoscopic instruments, adjust the height at which they stand, etc. They performed a basic laparoscopic bead transfer task or other similar task on the ProMIS simulator.

3.2.4.2 Simulated HALC cases

Subjects performed five simulated HALC cases on the ProMIS simulator by completing a series of predetermined, standardised procedural steps:

1. Tilting Bodyform
2. Identification of the inferior mesenteric artery (IMA)
3. Identification of the left ureter
4. Transection of the inferior mesenteric vein (IMV)

5. Mobilisation of sigmoid colon
6. Mobilisation of descending colon
7. Take down of splenic flexure
8. Creation of anastomosis
9. Testing of anastomosis

The initial 3 procedures were performed in “Teach Me” mode in which the simulator provides assistance by a virtual mentor (i.e., simulator provides instructions and annotations super-imposed on a video monitor) and augmented reality (i.e., simulator combines video and interactive environment/graphical objects). The subsequent 2 procedures were performed in “Test Me” mode; in which the simulator provides no simulator assistance.

All subjects were proctored by a facilitator. They received coaching on the first HALC procedure and for subsequent procedures they could ask questions. However, they were not prevented from committing errors.

3.2.4.3 Feedback – groups

Subjects were assigned to one of two groups for training purposes:

Group 1 – control (n=16)

Subjects in this group could ask questions but did not receive structured feedback after every session. They had different facilitators who did not give them standardised feedback.

Group 2 - feedback (n=12)

Subjects in this group assessed their own performance after every case. They then had the chance to review their own anatomy tray after every procedure with the facilitator and any errors made were pointed out. Where the subject was unaware that an error

had been committed the relevant part of the tray was examined and the error explained. This standardised feedback was provided after every procedure. Instrument metrics (ie time, IPL and IS) were reviewed after every other procedure. All subjects in this group were assisted by the same facilitator. Subjects in group 2 also gave their opinion regarding simulator training before and after the session.

3.2.5 Formal assessment:

Each subject's performance was assessed from two sources

1. Simulator metrics

- Instrument Path Length (IPL)
- Instrument Smoothness (IS)

Both of these metrics are objectively measured and recorded by the simulator.

2. Anatomy tray assessment.

The anatomy tray from each procedure was uniquely identified and randomly allocated for independent, blinded, review by two evaluators trained to identify the occurrence of 14 predetermined surgical errors (Appendix 3.2). These errors have been demonstrated to be construct valid in a previous study (Neary PC et al, 2008).

Proficiency levels for the HALC procedure have been previously established by the authors in a separate study and were derived from the performance of 16 expert colorectal surgeons. They were not actively utilised in this study as attainment of proficiency was not the aim, but are listed for reference in Appendix 3.3)

3.2.6 Sub study – skills transfer assessment in Group 2

Subjects in Group 2 only also performed some basic simulator procedures following their HALC procedures. This was done in order to assess transfer of skills. As part of their “warm-up session” before they commenced the HALC cases, these subjects performed a virtual colonoscopy procedure on the GI Mentor Simulator, and virtual laparoscopic cholecystectomy procedure on the LapSIM Simulator, as described in Chapter 2. Group 1 subjects also had a “ warm-up” session but their data were not recorded. After completion of the HALC training, subjects in Group 2 repeated the same basic procedures twice. Performance data consisted of the metrics which were measured by the simulator.

3.2.7 Statistical analysis

The performance data for the two groups of subjects were analysed descriptively using SPSS software version 15.0 (SPSS Inc, Chicago, IL, USA) Mann-Whitney-U tests were conducted to determine whether the means for the instrument path length (IPL), instrument smoothness (IS), and error scores were equivalent across Group 1 and Group 2 subjects. The IPL and IS were measured five times each, and the first IPL measurements for the one sample were compared to the first measurements for the other, and so on for the other five cases. Although the data is numerical and continuous, as each trial was compared separately the sample size was not very large and histograms created from the instrument data did not lie under a bell shaped curve, Therefore it was felt non-parametric analysis of the data was more appropriate.

3.3 Results

3.3.1 Demographics

All subjects completed all 5 trials. There was no significant difference between the two groups for age (mean age 31.5 vs 32.27 years), hand dominance or previous experience of advanced laparoscopic, laparoscopic colectomy, or hand-assisted laparoscopic procedures.

3.3.2 Overall results, learning curve

When the results of all 28 subjects were analysed together, there was a significant learning curve observed in the instrument scores across the 5 trials. (MANOVA, IPL $p = 0.000$, IS $p = 0.000$, figures 3.1 – 3.3)

Figure 3.1 - Mean results for IPL for all 28 subjects

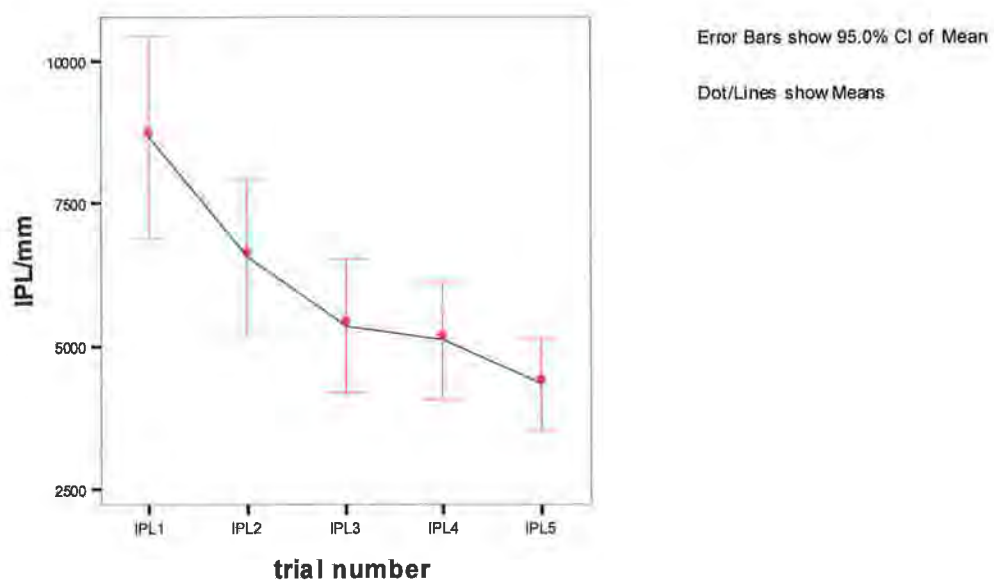
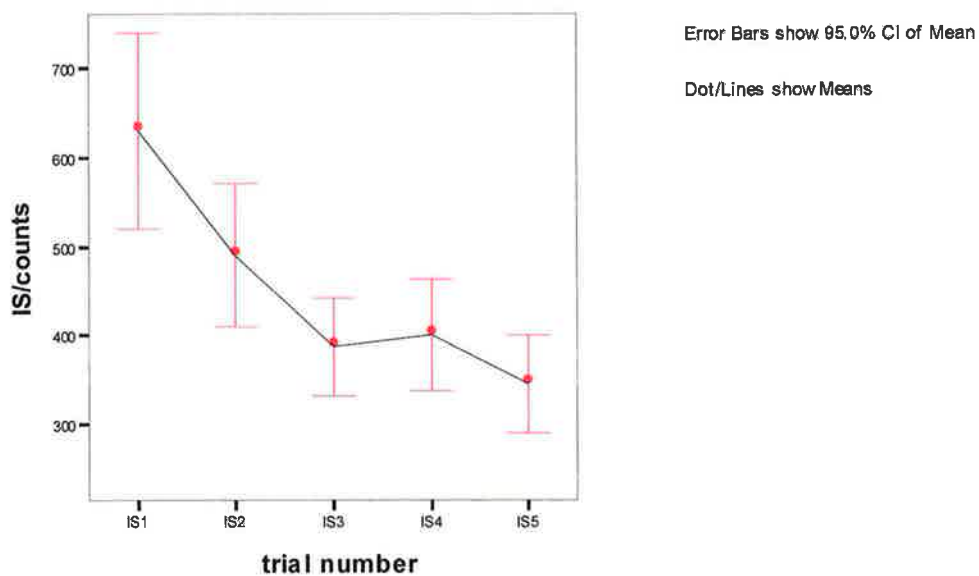
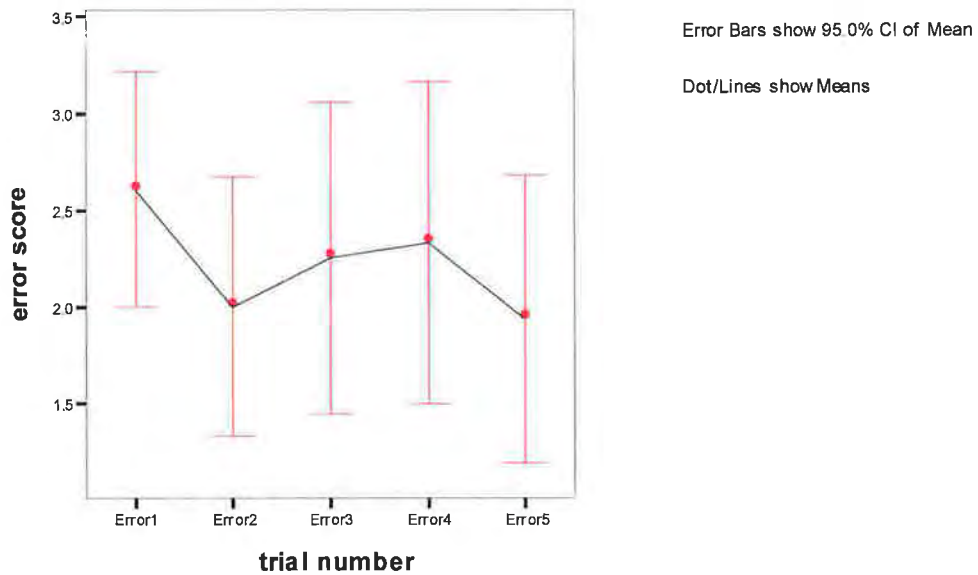


Figure 3.2 – Mean results for IS for all 28 subjects



When error scores were analysed for the entire group, a significant improvement was not seen. (ANOVA for repeated measures, $p = 0.661$)

Figure 3.3 – Mean error score for all 28 subjects



3.3.3 Results for groups 1 and 2 - effect of feedback

Overall results for groups 1 and 2 are displayed in table 3.1.

Table 3.1

	Group 1 - control	Group 2 - feedback	P values (Mann-Whitney-U)
IPL mean (st dev)/mm	23, 874 (8999)	39,086 (11,042)	0.001
IS mean (st dev)/count	2015 (632)	2567 (690)	0.045
errors	14 (5.44)	4.42 (2.57)	0.000

Results for IPL for each group separately groups are displayed in Figure 3.4, and the IS results in Figure 3.5. As can be seen from the table and figures, the instrument scores (ie IPL and IS) for Group 1 improved across the 5 trials with a slight decrease on trial 4. In addition, their performance as a group became more consistent over the 5

trials as shown by the decreasing standard deviations from trial 1 to 5. Similarly the instrument results for Group 2 show an improvement across the 5 trials, but with no decrease on trial 4. Their performance also became more consistent. However, when the results from both groups are compared, it is clear that Group 2 performed worse for IPL ($p = 0.001$, Mann-Whitney-U), and worse for IS. ($p = 0.046$, Mann-Whitney-U). Looking at each trial separately, for IPL, Group 2 performed significantly worse on trials 2, 3 and 5, and for IS, Group 2 performed significantly worse only on trial 3 and trial 5. This occurred despite the feedback given to Group 2 which was predicted to improve performance. The subjects in group 1 reached proficiency for their instrument movement scores on their last trial, but the subjects in group 2 did not (Appendix 3.3)

Figure 3.4 - mean results for IPL for groups 1 and 2

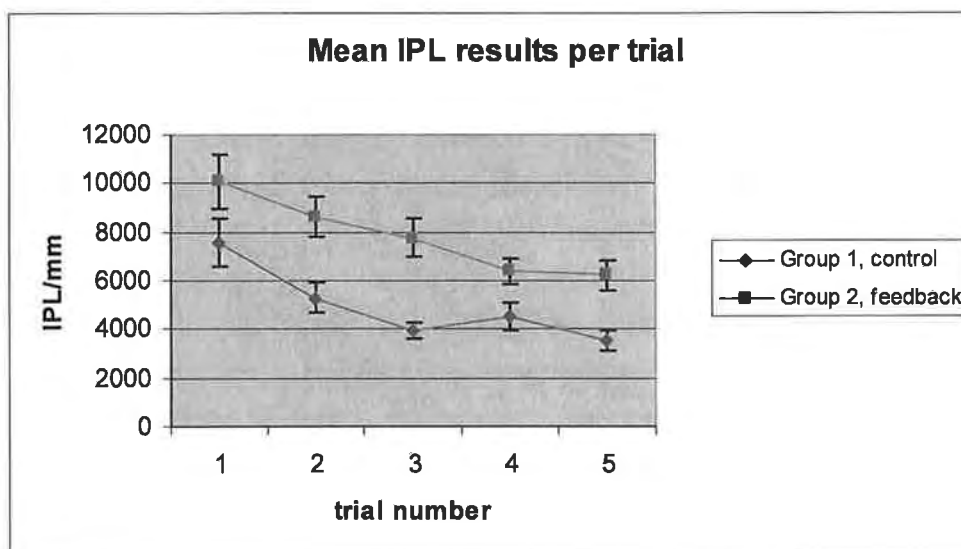
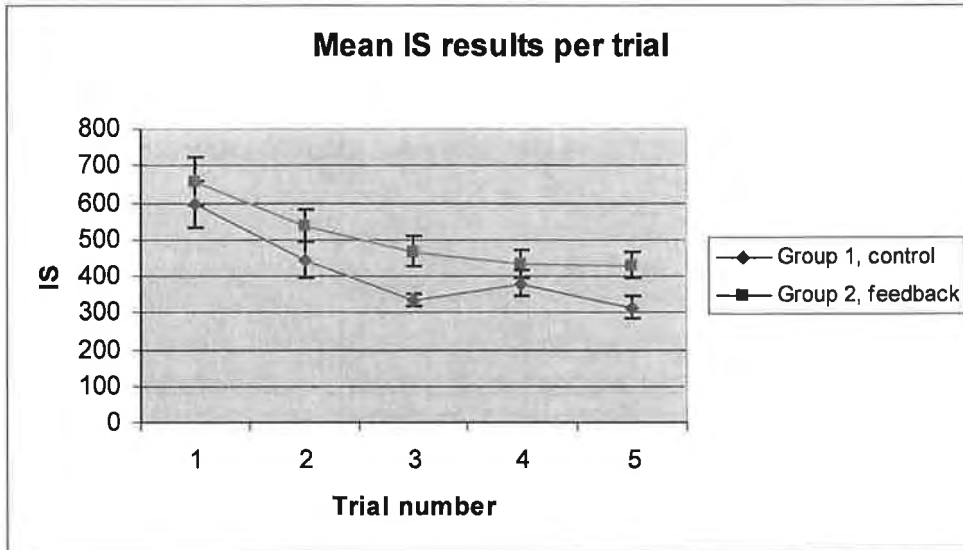


Figure 3.5 – mean results for IS for groups 1 and 2



Error scores revealed a different pattern. Mean error scores across the 5 trials in groups 1 and 2 are presented in Table 3.1 and Figure 3.6. Group 2 made significantly less errors overall ($p = 0.000$, Mann-Whitney-U) and when the 5 trials are examined separately, the subjects in group 2 made significantly less errors on all trials except trial 2. In addition as can be seen from the figure, they had a smoother learning curve a demonstrated a plateau by trial 4. No such plateau was evident in group 1, nor was there a clear learning curve.

Figure 3.6 – mean error scores for groups 1 and 2

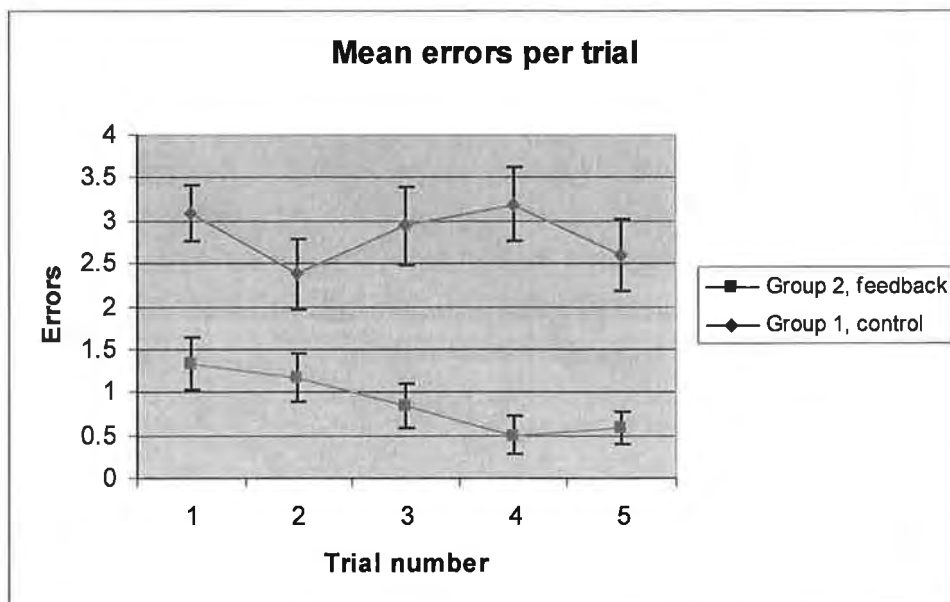


Table 3.2 – Mann–Whitney–U test results for significance between the two group’s mean performance results

procedure	1	2	3	4	5
IPL	0.071	0.003	0.002	0.064	0.003
IS	0.283	0.071	0.021	0.407	0.002
errors	0.004	0.094	0.003	0.000	0.000

Inter parameter correlations using Spearman’s rho correlation coefficient support this finding, ie a poor correlation between the instrument efficiency scores and error scores. (Table 3.3) The correlation between IPL and error scores was weak and negative, the correlation between IS and error scores was similar and was non significant. However, as expected the correlation between the 2 measures of instrument efficiency ie IPL and IS was significant (0.906, $p = 0.001$)

Table 3.3 – inter-metric correlation

	Spearman’s rho	Significance (2 tailed)
IPL and error scores	- 0.470	$p = 0.013$
IS and error scores	-0.272	$p = 0.169$
IPL and IS	0.906	$p < 0.001$

3.3.4 Self-assessment and subject opinion

As part of the structured feedback process, subjects in group 2 had to indicate errors they thought they had made before reviewing their anatomy tray. They had to recall these errors without the marking sheet, so many subjects did not accurately remember which structured errors they had committed. They also assessed their overall performance using a Leikert 5-point scale as used in the OSATS assessment system. (Appendix 3.4, Reznick et al, 1997). To assess the accuracy of self-assessment, we correlated IPL scores with the OSATS score for

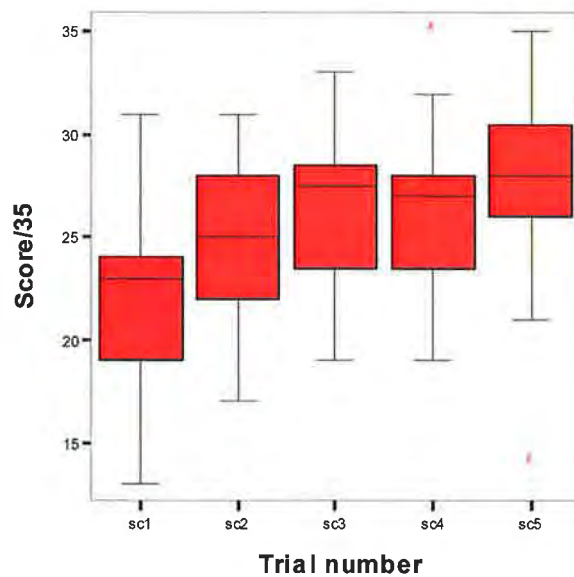
all 60 procedures performed by Group 2 using a non-parametric correlation coefficient, and found a significant but weak correlation of -0.567 , $p = 0.000$. Table 3.4 shows individual correlations for each subject, (data missing for 2 subjects). None of the results reached statistical significance, likely due to the small number of 5 analyses per subject.

Table 3.4 – individual subject correlations between IPL and global self-assessment

Subject	Correlation co-efficient
1	-0.800
2	-0.718
3	-0.154
4	-0.783
5	-0.319
6	-0.821
7	-0.316
8	-0.462
10	0.359
12	-0.031

As can be seen from figure 3.4 and 3.7 the self-assessment scores reflected the actual scores achieved by each subject

Figure 3.7 – global self-assessment scores per procedure for each subject in group 2



We also assessed the opinion of the subjects in Group 2 relating to surgical simulation. (Appendix 3.1) They were asked to score their opinion using a 1-5 scale, where 1= disagree strongly and 5 = agree strongly. The scores were averaged for each question and are presented in Table 3.5, a,b.

Table 3.5 a – Pre and post training opinion scores for group2. Each score is mean of subject’s individual scores out of 5

Question	Responses pre-training	Responses post-training
Simulator-based training is useful in surgical training	4.2	4.4
Training on a simulator helps improves real surgical performance	3.9	4
Simulator training should be obligatory for all trainees	3.8	4.2
In your opinion, will simulator training become obligatory for all trainees	4.5	4.4
Simulator-based training will improve patient safety	3.9	4
Will this session be beneficial for you?	4.3	na

Table 3.5 b – Post training opinion scores for group2. Each score is mean of subjects' individual scores out of 5

Question	score
Was today beneficial for you?	4.7
Do you think this training will improve your operative skills?	4.4
Would you feel more confident about performing laparoscopic colorectal surgery after today?	4.3
Would you feel more confident about performing laparoscopic surgery in general after today?	3.9
Would you be happy to attend a similar course again?	4.7

3.3.5 Inter rater reliability

The instrument efficiency scores are objectively calculated by the simulator but the error scores were “manually” graded by two assessors blinded to the subject identity. A detailed set of error definitions were used as described in chapter 2.

Two assessors were used in order to increase the robustness and objectivity of the scoring system. Trays were scored independently, where a difference in score existed the mean score from the two raters was used, and where inter rater reliability (IRR) fell below 0.8 for any individual tray, the score was reviewed by both assessors and a consensus was reached. In all cases, this was due to assessor error, and the score was changed accordingly. IRR was calculated using the formula

$$\text{IRR} = \text{number of agreements} / \text{number of possible agreements}$$

In some cases, both reviewers may have given the same score for a tray ie 12/14, but may have graded different errors. For example, assessor 1 may have identified error 1 and error 6 on a tray and assessor 2 may have identified error 1 and error 7. Although both graded the tray as 12/14, the IRR would be 0.86, as they had two disagreements.

The IRR for all 140 anatomy trays was 0.969 (sd = 0.054) and fell below 0.8 for only 3 individual anatomy trays

3.3.6 Skills Transfer Results

Several metrics are generated by the simulators for the virtual laparoscopic cholecystectomy procedure and colonoscopy.

The 12 subjects in Group 2 each performed both procedures 3 times – once before the HALC training and twice after. Demonstration of skills transfer would likely have involved a larger improvement between attempts 2 and 3 than between the other attempts.

The mean results for colonoscopy are listed in Table 3.6, and some of the results are displayed in figures 3.9, 3.10 As can be seen, none of the differences reached statistical significance although the mean scores for attempt 3 were all better than attempt 1.

Table 3.6 – mean colonoscopy results for attempts 1-3, with p-values calculated using Kruskal-Wallis tests

metric	1	2	3	p-value
% mucosal surface examined	77.8	67.2	69.7	0.248
time to reach cecum	489	467	329	0.382
%efficiency of screening	73.77	77.5	77.22	0.258
%time clear view lumen	90.2	85.43	90.8	0.537
excess local pressure	1.789	2.25	1.43	0.835

Figure 3.9 – results for % mucosal surface examined, efficiency of screening and time with clear view over the three attempts

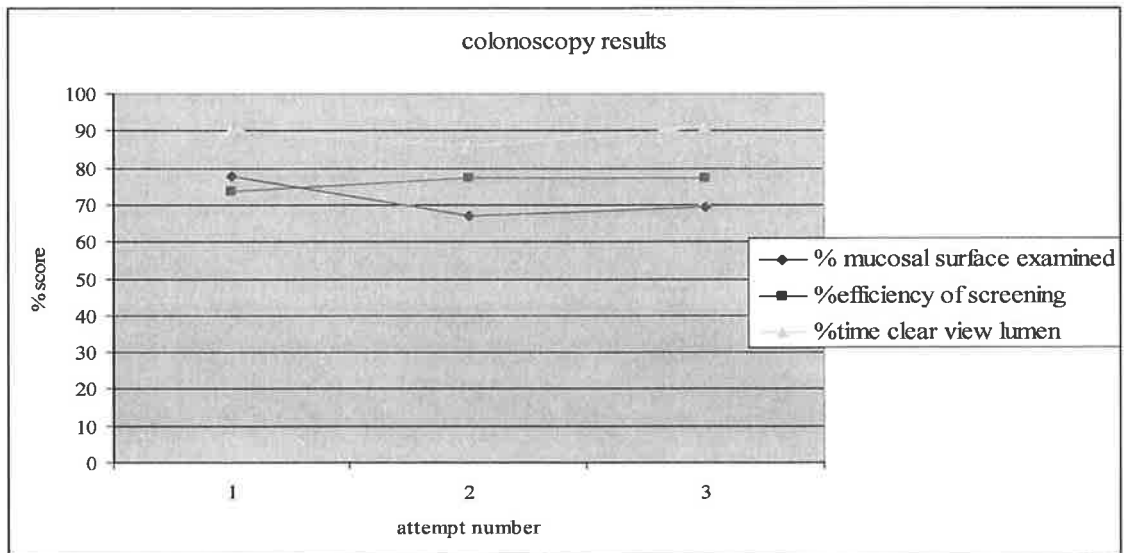
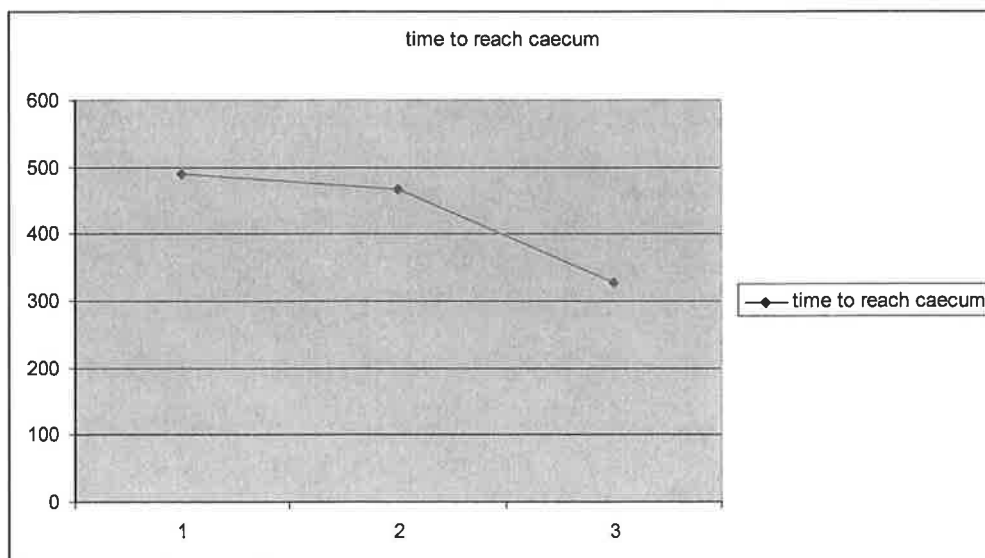


Figure 3.10. Mean results for time to reach the caecum

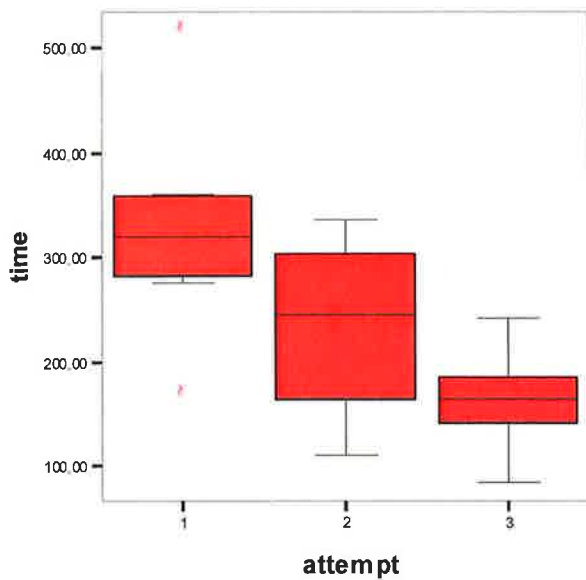


A large number of performance outcomes were also generated for the laparoscopic cholecystectomy task. (Part 2 of the task, removal of gallbladder from liver bed was analysed as the results for part 1 were unsuitable) Several of these showed significant differences when analysed together (Table 3.7), but when within-subject differences were analysed using ANOVA for repeated measures, only the metrics for time and angular path length differed significantly across the 3 attempts ($p = 0.010$ and $p = 0.036$ respectively)

Table 3.7 – metrics for the laparoscopic cholecystectomy task which showed significant differences (Kruskal-Wallis).

Metric	P- value
Time	0.003
Energy burnt in air	0.076
Left angular path length	0.021
Instrument path length	0.011

Figure 3.11 – results for time to complete the laparoscopic cholecystectomy task across the three attempts



3.4 Discussion

As expected, subjects in the study showed a gradual improvement across the 5 cases in terms of instrument use. They performed better and more consistently with regard to instrument use scores from their first to their fifth trial. This has been shown consistently in the literature, ie that repeating, or practising a task leads to an improved performance (Sedlack RE et al 2002; Ferlitsch A 2002; Neequaye SK et al, 2007). This is one of the advantages of simulation as subjects can repeat a task until they reach a desired performance standard. The subjects in the control group did not improve from trial 3 to trial 4. This may have been due to the change in the simulator mode from “teach me” to “test me”. This pattern was not seen in the feedback group, possibly indicating that the structured feedback they were receiving made them less reliant on the teaching aids provided by the simulator.

Although none of the subjects had performed a HALC procedure before the study, all subjects were able to perform complete procedures from their first trial. This demonstrates partly the value of the cognitive skills, or diadactic part of the curriculum (Van Herzeele I et al, 2008) ; a simulator can train but not teach. Subjects should be familiar with the procedural steps including common errors, and other information such as relevant anatomy and physiology before attempting a procedure on a simulator in order to obtain maximum benefit. As demonstrated in the study, simulators are most effective when incorporated into structured teaching curricula (Gallagher AG, 2005)

We hypothesised that providing structured proximate feedback would improve performance. Error scores were significantly improved in the feedback group but at the expense of IPL. This is not surprising when the nature of the anatomical tray is taken into account. For nearly every part of the procedure, performing a more careful dissection necessitates more instrument usage, thereby augmenting these scores. For example, dividing the IMA properly requires it to be completely dissected out from the mesentery surrounding it, otherwise the stapler will apply staples across the mesentery which counts as an error (error 3). Performing more dissection however requires more instrument movement. A similar issue arises with steps 6 and 7, which involve dividing the lateral colon from its attachments. Again, performing this step correctly requires the trainee to divide three layers of tissue. Frequently a subject will perform a superficial dissection, realise it is not sufficient, and repeat the step. Clinically this is appropriate behaviour but it leads to higher instrument movement scores. This was the case in this study, where the subjects in group 2 had better error scores, but worse instrument movement scores than their counterparts in group 1.

These data raise interesting questions about the metrics used to analyse simulated surgical procedures. IPL and IS metrics are well validated, and used in a variety of other simulators in addition to the ProMIS. They have been proven to demonstrate construct validity in many studies, not only between experts and novices (Neary PC et al, 2008) but also between intermediate standards of surgical experience (Pellen MG et al, 2009). They and similar metrics have been utilised in gynaecological and arthroscopic simulators (Tashiro Y et al, 2008). For example, in the paper by Tashiro et al, probe path length and probe velocity are measured, as is force of instrument use and all these metrics were found to distinguish surgeons of different experience levels. However there is some evidence to suggest this is not an optimum way to measure performance (Chmarra MK et al, 2007), and little research comparing these metrics to other performance measures. What little there is seems to focus on more basic tasks such as the peg transfer task in the fundamentals of laparoscopic surgery programme (FLS) (Ritter EM et al, 2007), or the sharp dissection task on the ProMIS (Pellen M, 2009). These studies have shown a correlation between IPL and IS and outcome measures on these tasks. With other more clinically relevant tasks there is some evidence that these metrics may not correlate with more clinically relevant outcomes (Cesaneck Pet al, 2008) The more advanced tasks used in the study by Cesaneck et al included vessel clipping and dividing, loop appendectomy and mesh placement with tacks. There was no correlation between instrument path length and smoothness and clinically relevant outcome measures such as knot slippage or tissue damage. Our results are similar in this regard. We feel performing a safe, thorough and error free procedure is more desirable than achieving efficient instrument movement and postulate that with more experienced surgeons, or more training on the simulator there may be a stronger correlation between the two performance measures. Given the

nature and design of the tray, it is intuitive that performing a more thorough tray dissection will augment movement scores in a novice cohort. This appeared to affect path length marginally more than smoothness, which differed significantly on fewer trials.

Achieving proficiency is the goal of any training programme. If providing structured feedback helps trainees to achieve proficiency in a shorter time this has implications for length of training in an environment where training hours are already severely limited. In addition it will help to reduce the costs of training which are considerable given the consumables required in an augmented reality simulator. The ideal next step in a study such as this would be to demonstrate an improved performance in the clinical environment. However there is much evidence to suggest that skills acquired in the virtual environment will transfer to the real operating room if proficiency has been achieved (Seymour N et al, 2002; Grantcharov TS et al, 2004) In particular, the improved error scores are relevant to clinical practice, and should lead to trainees operating in a safer manner in real life.

The value of feedback has also been demonstrated in the clinical setting (Harewood GC et al), although it would be difficult to give such immediate and objective feedback as in this study. The advantage of the simulator is being able to point out to trainees their errors after every procedure, or allowing them to assess their own errors. They can examine their anatomy tray immediately after each procedure and review it again if necessary. Without this feedback, a trainee may be unaware that they had committed a particular error and they may have continued to make the same mistake, allowing the simulator to reinforce undesirable behaviour. Although we demonstrated that standardised structured feedback improved performance, this feedback could be provided by non surgical faculty, provided the facilitator is sufficiently familiar with

the procedure. With regard to the anatomy tray, the errors are clearly defined in non-clinical terms. Minimal extents of dissection for example are confirmed using ruler measurements from specific points or structures on the tray, ie the correct extent of division of the splenic attachments are verified by using a simple ruler to measure the distance from the inferior tip of the spleen. This does not require clinical experience, Similarly the simulator metrics are explicitly provided and do not require clinical interpretation. An added benefit of the simulator, which was not utilised in this study, is the ability to observe video playback of each procedure as all procedures are recorded by the simulator. This could also be used to give feedback to trainees, and might potentially have an effect on instrument movement scores as the trainee could objectively observe their instrument movements after completing a procedure.

The provision of performance-related feedback should be structured and standardised. In our study, subjects in Group 2 had to self-assess their own performance and reflect on their error scores before reviewing the anatomy trays with the facilitator. This is a valuable part of the feedback process and enhances the learning experience. We did not directly compare self-assessed with actual error scores as the numbers were small and subjects did not have access to the same scoring sheet as the assessors. Subjects also gave scored their performance using a global scale. This score was correlated with IPL only as this was closely related to IS (Table 3.3) Overall this was significant but weak (- 0.567) but when the scores for individual subjects were compared there was a wide range, with some subjects' self-assessment scores correlating very highly with actual scores (as high as 0.8, Table 3.4). Other subjects had lower correlation coefficients and one subject had an inverse correlation. Therefore it is clear that some trainees can more accurately self-assess then others.

The training session was enjoyed by all subjects. Although all subjects had high opinions of simulation training before commencing, all subjects improved their opinion rating following the training. Subjects agreed strongly that they had benefitted from the session and would be willing to re attend a similar training session.

One criticism of the study is that although Group 2 made fewer errors than Group 1, their result for their first trial was also better, so their superior performance could potentially be due to a more advanced baseline performance and not to the provision of feedback. However both groups were similar in terms of previous HALC and laparoscopic colectomy experience. Group 2 had more standardised coaching for their first trial which may have improved their first performance, but the shape of the learning curve over the 5 trials (figure 3) shows clearly that Group 1 subjects did not improve throughout the course of the procedures. Had they done so they would have shown an even more impressive learning curve than Group 2. In addition, although all subjects in both groups had the same teaching before commencing, the feedback subjects were trained in smaller groups. This may have increased the knowledge transfer during the teaching session and may partly explain their superior performance on their first procedure.

We performed a small sub-study to analyse skills transfer in Group 2 only. The intensive training the subjects received in the HALC procedure would be extremely beneficial if we could demonstrate that the skills obtained would transfer to other procedures. This would increase the value of skills training for one particular procedure. There was no evidence of skills transfer to the colonoscopy task as we found no significant differences between the subjects' performance of this task before and after the HALC training (table 3.5 and figures 3.7 and 3.8) Although both

procedures are image guided and involve manipulation of a device at a distance from the operating field, it could be argued that the procedures are too dissimilar to have skills overlap. There was some evidence of improvement in the laparoscopic task, which would involve a more similar skills set to the HALC procedure. Actual within-subject analysis showed a significant on only 2 metrics, however. Neequaye et al. (2007) demonstrated skills transfer between 2 procedures, but these were both endovascular procedures on the VIST simulator differing only in anatomical location and as such it is almost intuitive that there would be an element of skills transfer. It would be more useful to demonstrate this between different procedures such as was attempted here. However, it is important to consider that any improvements demonstrated do not necessarily imply causality, and that subjects in our may have improved simply because they repeated the other procedures 3 times, not because they performed 5 HALC cases inbetween. A final point is that as all testing was performed within one day, fatigue was a likely confounding factor which may have disguised any skill transfer which took place.

3.5 Conclusion

In conclusion the provision of standardised feedback during training was associated with significantly less errors and an improved learning curve. This occurred at the expense of instrument efficiency scores. Provision of structured, immediate performance feedback appears to optimise and improve the efficiency of surgical training and may improve patient safety. Trainees felt they had benefitted from the training session. Training to proficiency in one procedure may lead to improved performance in different, but related procedures.

Chapter 4

The effectiveness of endovascular simulator training for novices; can learning take place in the absence of expert feedback?

Chapter 4 The effectiveness of endovascular simulator training for novices; can learning take place in the absence of expert feedback?

4.1 Background

Over the past decade, therapeutic options for the treatment of vascular disease have changed greatly. In a similar fashion to the development of minimally invasive techniques for general surgery, the trend is now towards the development and refinement of catheter-based minimally invasive interventions for the treatment of vascular pathology. The advantages of endovascular procedures over open vascular procedures for the patient include a less invasive procedure with a shorter recovery time and reduced pain. Studies such as the SAPPHIRE trial have shown that carotid artery stenting (CAS) for example is safer than endarterectomy for the management of carotid artery disease in certain high risk patients (Yaday et al, 2004) and the FDA has approved its use in this cohort. The skill set required to perform these procedures differs from the skills required to perform open vascular surgery. The surgeon must be able to perform an image-guided procedure while using instruments and devices with limited degrees of freedom. It has been shown that there is a learning curve associated with these procedures, such as CAS where a clear correlation has been demonstrated between case numbers and complication rate (Lin PH et al, 2005) It is clear that interventionalists must be experienced before attempting these more complex procedures. This is said to be one of the reasons for the results of the EVA-3S trial which also compared CAS and endarterectomy, but had to be stopped early because of poor outcomes in the CAS group. In this study, CAS was performed by relatively

inexperienced interventionalists (Mas JL et al, 2006). Current vascular surgery training does not adequately prepare surgeons to perform these complex procedures. In addition, interventionalists from fields such as radiology, cardiology and neurosurgery are now performing these procedures. This variation in speciality brings different expert skills sets to the table, ie a vascular surgeon will have a thorough understanding of vascular anatomy and pathology and a radiologist will have extensive experience with wire and catheter manipulation and management of the fluoroscope. However, with such different backgrounds the establishment of endovascular training programmes becomes even more important, in order to standardise the competency of endovascular interventionalists (Gallagher AG et al, 2004).

Despite the high-risk and complex nature of these procedures, one of the major components of training is the supervised performance of these procedures on patients. As logistical and ethical factors are limiting this apprenticeship training, it is intuitive to supplant the early part of the learning curve with simulator training in a risk-free environment. Virtual reality simulator training in particular appears to be well suited to endovascular skills training, as it can easily mimic the real life situation of manipulating a wire or catheter in a real, 3D field, while viewing this activity on a 2D monitor. Although there are other training platforms available for vascular surgery (Neequaye et al, 2007), the use of high fidelity VR simulators enables trainees to learn basic wire and catheter handling skill in a safe and economical environment, and affords expert practitioners the opportunity to refine and refresh procedural skills. There is the added advantage of mission rehearsal for more experienced practitioners, whereby actual patient data can be entered into the simulator, allowing the interventionalists to practice the actual procedure they will then perform.

Compared to other areas in surgical training, relatively less evidence exists to prove the value of VR simulation for endovascular procedures. While a number of studies have been carried out which demonstrate face, content and construct validity of VR simulators in this area, there is only one study which demonstrated skills transfer to human patients (Chaer RA, 2006). However, the educational value of high fidelity simulators is not debated. Many training studies have demonstrated a learning curve with repeated practice on a simulator, which tends to benefit the novice in particular as the expert learning curve is much shorter (Gallagher AG et al, 2004). Trainees can practice a standardised procedure until they reach a certain benchmark level. This is useful for curriculum development, as trainees would have to attain this level before they would be allowed to perform any procedure on a patient.

In 2004, VR simulation training was approved by the FDA in the United States for carotid artery stenting training in recognition of mounting evidence that this training can improve intraoperative performance (<http://www.fda.gov> 2004, Gallagher et al, 2004)

Given the expense and sensitivity of these simulators, they are commonly used for training in intense skills courses, and this type of training has been shown to improve performance in experts (Van Herzeele I, 2008). These courses are expensive to run, particularly when expert faculty are involved, and the objective is to maximise the learning potential of these courses for every participant. It has been shown that a cognitive teaching component can enhance the learning experience for trainees (Van Herzeele I et al, 2008 (2)) and this is a feature of most curricula involving simulation. A simulator itself is not a “stand alone” educator, but rather is most effective when integrated into a teaching curriculum, as it has been said that simulators train but do not educate.

As discussed already in Chapter 3, another factor which has been shown to be important for effective learning is the use of feedback. Mahmoud et al demonstrated that there was no learning curve on a colonoscopy simulator in the absence of feedback (Mahmoud T, 2004). A range of other studies have shown improvement in outcome parameters and more efficient learning when feedback is provided, although there is no consensus regarding the optimal style of feedback (Sewell C et al, 2008, Stefanidis et al 2007, Van Sickle K et al, 2007) For example one study which trained subjects on a suturing task suggested that simply informing the trainees of their results after each trial may be as effective as expert instruction (O' Connor et a, 2008). As there is a high cost and logistical difficulty associated with providing expert faculty for surgical training courses, it would be advantageous to know if this is necessary for optimal learning to take place. We propose to investigate this by comparing the effect of expert feedback, non-expert feedback, and no feedback on trainee learning curves for a simulated endovascular procedure.

Primary aim

To assess the need for expert instruction for novices learning catheter wire –based procedures on the VIST simulator.

Secondary aims

To shorten the novice learning curve for a catheter-guided procedure on the VIST simulator.

To examine performance results generated by the VIST simulator and those generated by objective assessment and self-assessment.

4.2 Materials and Methods

4.2.1 Ethical approval

Ethical approval was obtained from the Research and Ethics Committee RCSI prior to commencement of the study.

4.2.2 Setting

The study took place in the surgical skills training lab in the National Surgical Training Centre at RCSI.

4.2.3 Simulator

The simulator used in this study is the ProCedicus VIST™ (Vascular Interventional Surgical Trainer) system (Mentice, Gothenberg, Sweden) as described in Chapter 2.

4.2.3 Participants

The study population comprised surgical trainees who are novice in endovascular procedures.

Inclusion criteria

Completed Basic Surgical Training (BST)

Exclusion criteria

Commenced Higher Surgical Training

Previously performed any complete or partial endovascular procedures

Prior experience on VIST simulator, or any other high-fidelity vascular simulator (Observation or assisting at an endovascular procedure was not considered an exclusion criterion)

Facilitators

Two categories of facilitators were involved in the study, expert facilitators, and a non-expert facilitator. All facilitators underwent a standardised training programme for the VIST simulator, and were familiar with the simulated procedures and metrics.

Expert facilitators

Consultant vascular surgeons who perform endovascular procedures and are currently working in Ireland

Technical facilitator

Surgical trainee with no clinical vascular or endovascular experience.

4.2.4 Study procedures

18 trainees were recruited, and enrolled as subjects into the study on satisfaction of the inclusion and exclusion criteria. Demographic data were collected, including information regarding prior laparoscopic surgical experience, and video-game use. (Ability in simulators to selectively cannulate a target artery has been demonstrated to correlate with endovascular expertise though also with the number of videogame hours played per week (Dankleman et al, 2004) All candidates had the chance to read the subject information sheet and ask questions before signing a consent form.

4.2.4.1 Visuospatial testing

All subjects underwent a standard battery of perceptual, psychomotor and visuospatial tests as described for the colonoscopy study in Chapter 2. This combination of tests has been used at RCSI for several years for selection of candidates onto higher surgical training schemes. This was to ensure that no significant baseline differences existed between the groups which may influence performance. (Also to assess for any correlations between innate abilities and performance over the 6 trials, as discussed in chapter 7)

4.2.4.2 Didactic teaching

The procedure selected for the study was a renal artery angioplasty and stenting (RAS). This the second case in the renal artery module on the simulator. This was standardised for every case and was a left proximal renal artery lesion. The exact steps for performing this were taught in a standardised fashion to every subject. Subjects were initially required to perform an angiogram. Then they had to cannulate the renal artery using a diagnostic catheter and guidewire, and then they had to exchange the diagnostic catheter for a guide catheter over the guidewire. They were directed to position the guide catheter at the ostium of the renal artery to direct the placement of a balloon, then a stent over the lesion. Finally they were required to perform a completion angiogram.

The teaching session comprised the following elements:

Lecture accompanied by power point slides. This included

- an overview of renal artery disease including physiology, pathology, clinical presentation, treatment options
- introduction to the simulator
- endovascular instruments
- steps of the procedure including video clips
- errors to avoid

Demonstration of the procedure on the VIST simulator with commentary

Subjects watched a complete procedure, with commentary including an explanation of all the steps and discussion of performance measurements such as fluoro and contrast use. Advice was given regarding the avoidance of common errors. This is outlined in Appendix 4.1. This was standardised, and the same advice was given to all subjects.

Post-didactic questionnaire

After the didactic, all subjects took a short true/false questionnaire to ensure that they had understood the material. They were told the correct answers after completion if they did not score 100%.

4.2.4.3 Procedures

All subjects then proceeded to perform 6 procedures each. They could not perform more than two procedures without taking a break. Most subjects performed their six procedures within one day, but where this was not possible because of scheduling difficulties candidates performed all procedures within a 3 –day period.

As ability to memorise correctly the steps of the procedure was not specifically being tested, all candidates had access to a set of written instructions outlining the basic steps of the procedure. (Appendix 4.2)

Although the procedure was not performed with a sterile field, it was kept clinically realistic in other ways. For example, when changing instrument, subjects had to remove the instrument fully from the guidewire even when the actual instrument being used for the subsequent step was the same physical instrument. All subjects were offered assistance with changing of instruments. In addition the facilitator recorded cineloops and roadmaps where requested and performed C-arm positioning and instrument selection on the simulator as instructed by the subject.

Subjects performed all procedures with the C-arm in an antero-posterior direction and they were not given any instructions regarding alteration of the position.

Candidates were not given any coaching or direction during the procedure unless there was a risk of damaging the simulator ie if they attempted to insert an instrument other than what had been selected or if they were moving their instruments too fast. If they had questions they were directed to read the written instructions. However, if they were still unsure of what to do next they were given appropriate instructions, as the aim was not to test cognitive memory of the procedural steps. They were given no other coaching during the procedure, even when errors were committed.

Due to time and scheduling restrictions the decision was made to limit the time allowed for any one procedure to 40 minutes. This did not affect most subjects, but on several instances some subjects who were performing their first or second procedure could not complete it within the allowed time. No subject performing their third or greater procedure was limited by time.

All subjects filled out a brief self assessment form after every procedure, (Appendix 4.3) this consisted of a Leikert-type section covering issues such as ease of instrument use and safety of the procedure. Subjects were also asked to list errors they had made, and were asked if there was anything they would do differently on subsequent procedures.

4.2.4.4 Groups

Subjects were randomised into 3 groups:

Group A – control

Subjects in group performed their procedures with assistance/facilitation as described above. Although they were aware of the duration of their procedure they were given no other feedback regarding their performance.

Group B – non-expert feedback

The subjects in this group performed their procedures in an identical environment to those in group a, except that they were given feedback after every procedure. The simulator generates a lengthy procedural report, not every element of this was discussed in detail, but three main areas were discussed:

Procedural time, fluoro use and contrast use

Accuracy of balloon/stent placement

Handling errors as assessed by the simulator.

In addition, they were given feedback relating to the advice that had been supplied before the procedure (Appendix 4.1), and also feedback regarding any additional errors made.

Group C – expert feedback

The subjects in this group performed the same 6 procedures as the subjects in the other two groups. They had an identical teaching session. However, group C subjects were given feedback after every procedure by an expert, who also observed their performance. Experts were instructed to give whatever feedback they considered appropriate. In practice all the experts went through the simulator metrics with the subject and then gave additional feedback/

4.2.5 Performance Assessment

The performance assessment included two main elements – simulator-generated metrics and video-based performance assessment.

4.2.5.1 Simulator metrics

The VIST simulator objectively records performance parameters for every procedure performed.

These parameters can be divided into 3 categories:

General – this provides results for fluoroscopic and total procedure time, and volume of contrast use

Per lesion report - this provides measurements for the appropriateness of the size of balloon/stent and the accuracy of the balloon placement and stent deployment. It also includes some parameters which were not covered in the teaching session such as time

that balloon is inflated for, or balloon inflation pressure. Therefore a selection of relevant parameters was used in the assessment. These were:

Table 4.1 – per lesion report

Placement accuracy of balloon
Residual stenosis
% of lesion covered by balloon
Balloon vessel ratio
Placement accuracy of stent
Residual stenosis
% of lesion covered by stent
Stent vessel ratio

Handling errors – details for errors made while using instruments such as the catheter entering an inappropriate vessel or moving without guidewire support. A score is given every time an error is committed. Some of the metrics in this section were not relevant such as those relating to the 14' guidewire as only a .35' guidewire was used for the procedure.

(When the VIST simulator was designed, these errors were introduced based on the recommendations of consultant surgeons and interventional radiologists.) The errors that were used for assessment are listed below:

Table 4.2 - error metrics

Catheter scraping against vessel wall
Catheter moving without support of wire
Selective catheter scraping against vessel wall
Selective catheter moving without support of wire
Guide wire in small vessel
Guide wire entered suboptimal vessel
Catheter entered suboptimal vessel

4.2.5.2 Video-based assessment

In addition to these assessment parameters which are measured by the simulator, every performance was videotaped. Two views were recorded – the computer screen

showing the fluoroscopic view was recorded and a separate camera was used to record the instrument use outside the simulator. In each case, the trainee's face was not recorded, so that the videos could be assessed in a blinded fashion. The recordings were made in order to provide a separate source of assessment as except for the general measures, many of the other VIST performance metrics have been criticised for showing poor construct validity. (Neequaye et al, 2007) Some of the errors were deemed unfair by the experts, for example catheter scraping against wall of the vessel as it is difficult to avoid this error when cannulating the renal artery.

A standardised assessment form was created for the video assessments (appendix 4.4). This consisted of 2 parts - a detailed error scoring sheet and a Leikert-type global assessment section. The errors were scored every time they happened even when the same error was repeated and all were weighted equally and given a score of 1. The Leikert scale contained 7 dimensions and was marked on a continuous scale from 0 to 4, and anchored with explicit descriptions and the lower, middle and upper ends of the scale.

4.2.6 Statistical plan

Data were analysed using the Statistical Package for the Social Sciences version 15.0 (SPSS, Chicago, Illinois, USA) Data from the subjects' performance were analysed using non-parametrical tests. The independent variable was the group to which the subject was assigned, and the dependant variables were the metrics generated by the simulator. Differences between the groups' mean scores were compared for significance with Kruskal-Wallis testing, and pairs od Mann Whitney U tests were used to identify specific statistically significant differences between the groups. This

was done because although the data approximated a normal distribution when Shapiro Wilks tests were conducted the null hypothesis was not satisfied for all data points. Improvements between the groups were compared using repeated measures ANOVA where within subjects comparisons were the trial number and between subject comparisons were the group to which the subject belonged.

4.3. Results

Eighteen subjects in total participated. All subjects completed all 6 trials but for 1 subject the simulator crashed for two of the procedures so only four attempts were recorded. For the purposes of analysis, this subject's results for trial 4 were repeated for trials 5 and 6 which was a conservative approach, as there had been a continuous improvement thought out the previous 4 trials and it would be reasonable to presume that this trend would have continued.

12 of the 18 subjects completed all 6 trials within one day. For 6 of the subjects this was not possible due to scheduling restrictions and in these cases all trials were completed within a 3 day period.

4.3.1 Demographics

Demographic data are presented in table 4.3.

Table 4.3 – demographic data

Mean age/yr	30.27 (3.54)	
Mean years since graduation	5 (1.15)	
Wear glasses	Yes 10	No 8
Video games	Yes 4	No 14
Endovascular procedures	None 11 Assisted 7 Performed 0	

The mean age of the subjects was 30.27 years, and this did not differ significantly between the three groups. Four of the 18 subjects stated that they played video games, but for less than 2 ½ hours per week

Eight of the 18 subjects had previously rotated through a vascular rotation, but in all cases this was at senior house officer (SHO) level only. No subjects had performed any endovascular procedures prior to their participation in the study although 7 of the 18 subjects had observed or assisted at an endovascular procedure.

4.3.2 Overall results for time, fluoroscopic time and contrast use:

When all 18 subjects' results were analysed together, the most striking finding was the presence of a significant learning curve. All subjects improved their performance from their first attempt to their last. The most common pattern was that a subject would be unable to complete the procedure on the first attempt, but would subsequently be able to complete it within the 40 minute limit, and by the 5th or 6th trial all subjects were completing the case in much shorter times.

The largest improvement was noted for procedural time. (Figures 4.1 a-d, Table 4.4)

Figure 4.1a – overall results for procedural time for all 18 subjects

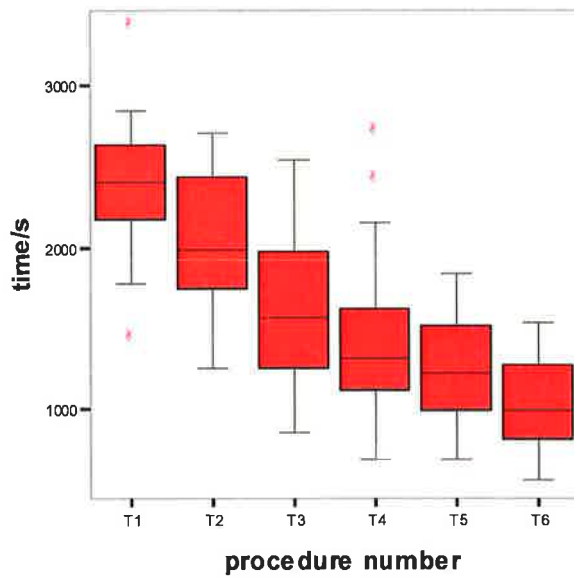


Figure 4.1b – overall results for fluoroscopic time for all 18 subjects

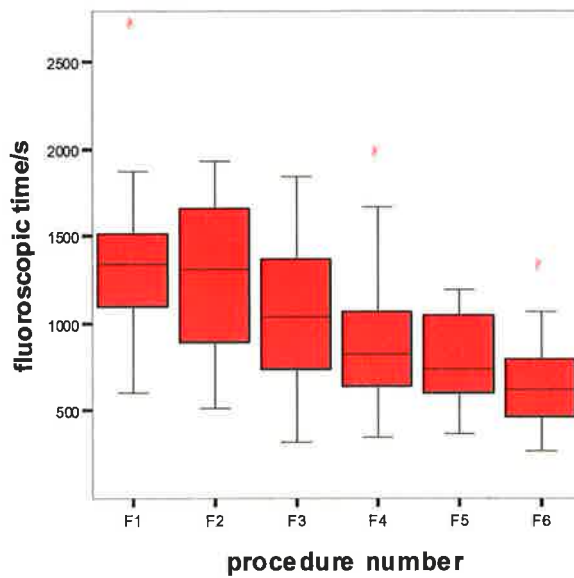


Figure 4.1c - Overall results for contrast volume for all 18 subjects

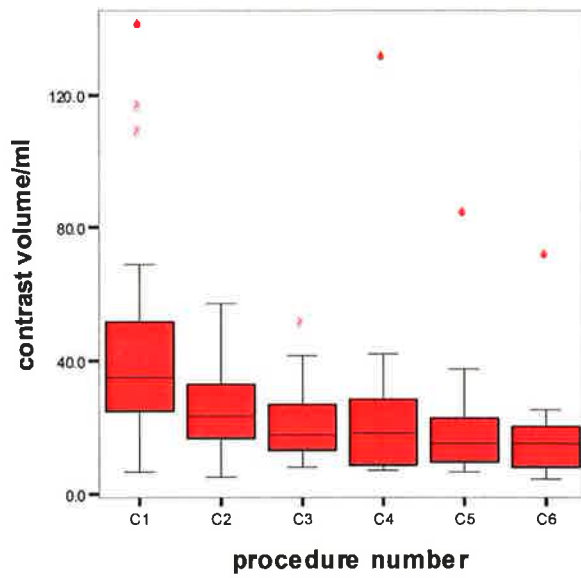


Figure 4.1d – Overall results for simulator error scores for all 18 subjects

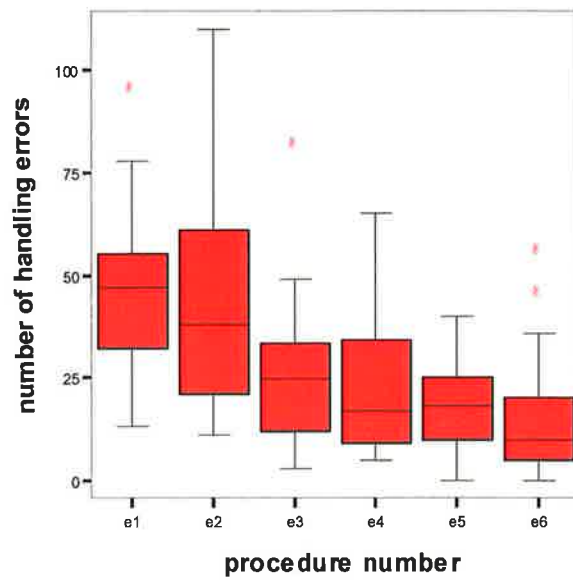


Table 4.4 – improvement between trial 1 and trial 6 for the simulator metrics

Metric	Significance (Friedman)
Time	p = 0.000
Contrast	p = 0.000
Fluoro	p = 0.000
Errors	p = 0.000
Placement accuracy	p = 0.047
Residual stenosis	p = 0.001
Lesion coverage	p = 0.000
Video errors	p = 0.000

4.3.3 Between group differences for simulator scores – effect of feedback.

Results for procedural time, contrast use, fluoroscopic time, error scores and accuracy of balloon placement scores were summed over the 6 trials for each subject. For each metric except placement accuracy, group C performed better than groups A and B, and for each metric except fluoroscopic time and residual stenosis, group B performed better than group A. However, when these results were compared using Kruskal-Wallis tests, significant differences existed between the groups for error scores only. Post-hoc testing revealed that the significant results were between groups A and C ($p = 0.009$) and groups B and C only ($p = 0.004$) (Table 4.5)

Further analysis of the errors scores by procedure number revealed that the differences between the groups were significant on trials 1, 4 and 5, although they approached significance on the other trials. (Table 4.6)

Table 4.5 – overall results for the 3 groups

metric	A	B	C	Significance (KW)
Time/s	10606	9809	8672	p = 0.104
contrast volume/ml	214	132	109	p = 0.148
fluoroscopic time/s	6212	6399	5255	p = 0.359
error scores	217	204	101	p = 0.009*
placement accuracy/mm	31	19	22	p = .296
residual stenosis	24	28	12	p = 0.333
lesion coverage/%	481	467	544	p = 0.584

Table 4.6 Significance of differences between the groups for error scores for each trial

Trial Number	p-value.
1	.037*
2	.092
3	.468
4	.031*
5	.011*
6	.052
total	.004

* significant

To further assess the effect of feedback, we compared the group's scores on their 6th attempt.

Table 4.7 – differences between the groups for trial 6

metric	A	B	C	p-value using Kruskal Wallis
Time/s	1014	1064	900	0.470
contrast volume/ml	24.9	17.2	9.55	0.078
fluoroscopic time/s	6212	6400	539	0.459
error scores	17	27	4.7	0.016*
placement accuracy/mm	5.18	2.87	0.85	0.019*
residual stenosis	19.5	13.5	1.5	0.791
lesion coverage/%	87.83	93.67	100	0.183

As can be seen from the results, group C out performed the other two groups in every metric, significantly so for placement accuracy and error scores and the difference for contrast approached statistical significance, and on a practical level the differences for this metric were marked (24.9 ml vs 17.2 ml vs 9.55 ml), table 4.7

4.3.4 Between group differences for video assessment scores

All videos were assessed in a blinded fashion.

For the error scores, the total errors for each group were averaged and are presented in Figure 4.2. As can be seen from the figure, group C performed better than group B, who performed better than group A. Significant differences were found between the groups ($p = 0.000$). Post-hoc testing using Bonferroni showed that the significant differences were between Group A, and Groups B and C. The differences between the global assessments were less marked (Figure 4.3). Figure 4.4 shows the mean results for each procedure in each of the 3 groups.

Figure 4.2 – mean errors for all 6 trials for each group.

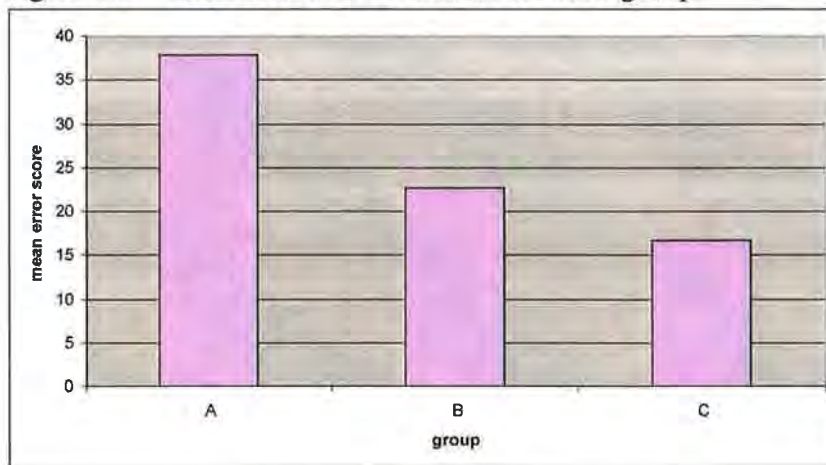


Table 4.8 – post hoc testing

Post-hoc comparisons	p- value
Grps A and B	0.002
Grps A and C	0.000
Grps B and C	0.428

Figure 4.3 – mean global scores for all procedures for each group

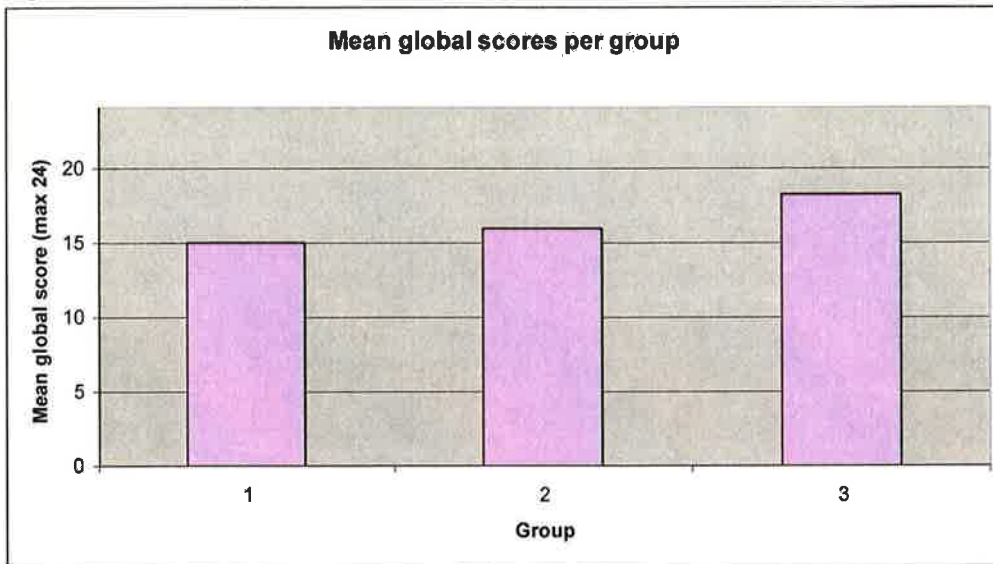


Figure 4.4a – mean video error scores per procedure

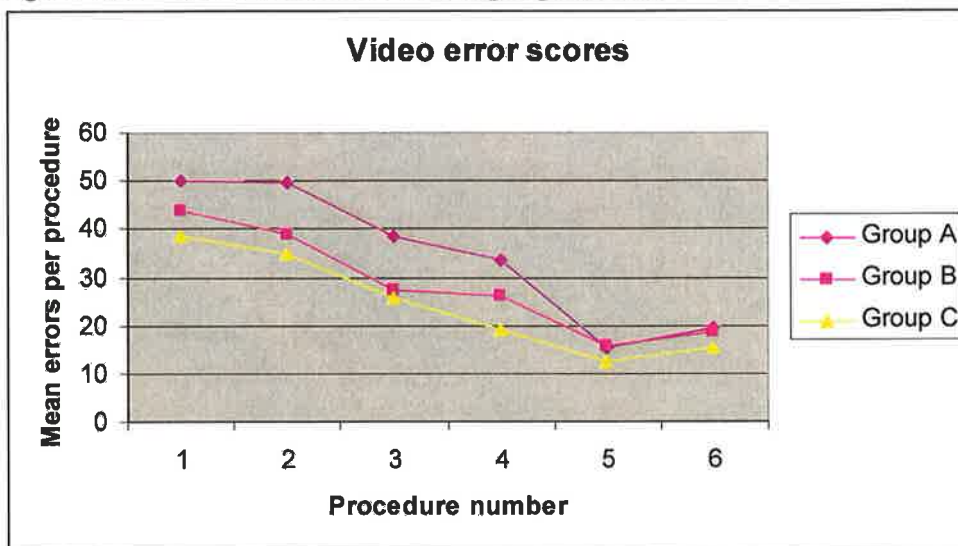
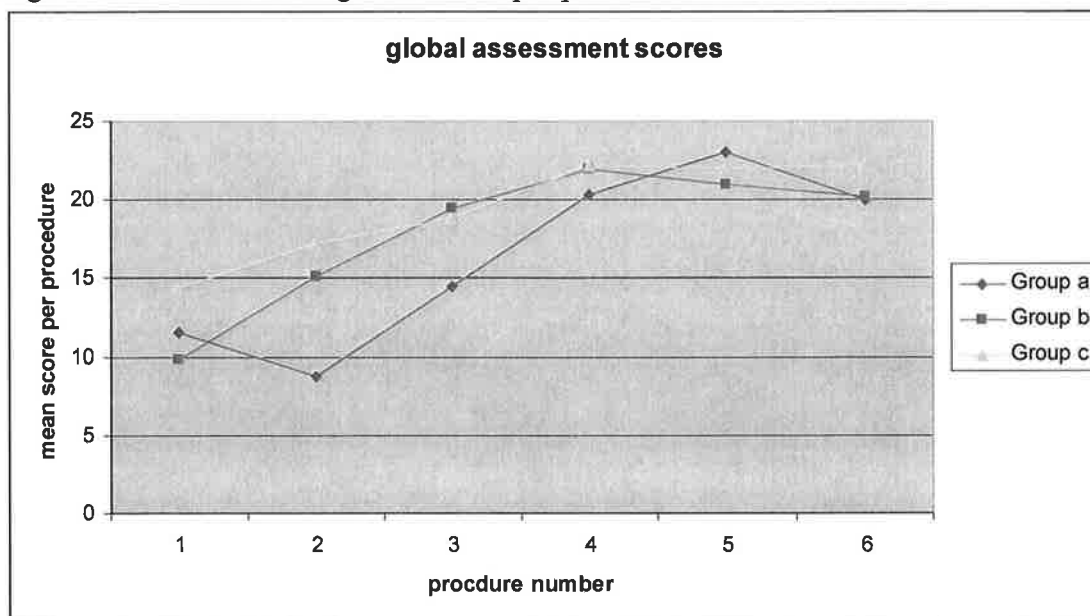


Figure 4.4b – mean video global scores per procedure



4.3.5 Effect of other factors

We investigated the effect of other demographic factors on the group's performance on their first and last trial, and found no significant differences between the results for the first and last trials when the group was divided by gender, wearing of glasses, previous endovascular experience or handedness. There was also no correlation found between their total scores for each metric and their general opinion of simulation as assessed on the pre-procedural questionnaire.

All subjects in addition had their fundamental abilities tested using the aptitude tests as described in chapter 2. There were 3 tests – visuospatial skills were assessed using paper tests from the Kit of Factor Referenced Tests, perceptual ability was assessed using the PicSO computer based test, and psychomotor ability was assessed using the location and coordination test on the ProMIS simulator.

There were no significant differences between the 3 groups for any of these results (table 4.9)

Table 4.9 – differences between baseline abilities for each group

Ability	Test	p-value
Visio-spatial	Average result for 3 tests (%)	0.366
Perceptual	PicSOr	0.490
Psychomotor	Time/mins	0.478
	Path Length/mm	0.208
	Economy	0.394

When results from these tests were correlated with each subject’s performance on their first trial, there was no significant correlation between any of the results except for one metric from the psychomotor test, path length. This correlated significantly with placement accuracy of the balloon, and also with 2 further related metrics – residual stenosis and %lesion coverage

Instrument path length correlations using Spearman’s Rho:

- Placement accuracy – 0.738, $p = 0.002$
- Residual stenosis – 0.640, $p = 0.014$
- % lesion coverage – 0.659, $p = 0.01$

4.3.6 Self –assessment

All subjects filled out a self-assessment form after each procedure, and those in groups B and C filled it out before they were given any feedback.

The form consisted of a 10- part questionnaire which allowed the trainees to give themselves a score between 1 and 5 for each of the 10 parts. (Appendix xx) It was constructed as a Leikert type continuous scale. They were also asked specific

questions about errors they had committed and what they would do differently on their next attempt.

We summed the scores for the 10 questions for each candidate, giving an overall score out of 50 for each trial.

The self assessment scores of all 18 subjects improved significantly over the 6 trials.

($p = 0.000$ using ANOVA for repeated measures) However there was no significant interaction with group as a factor.

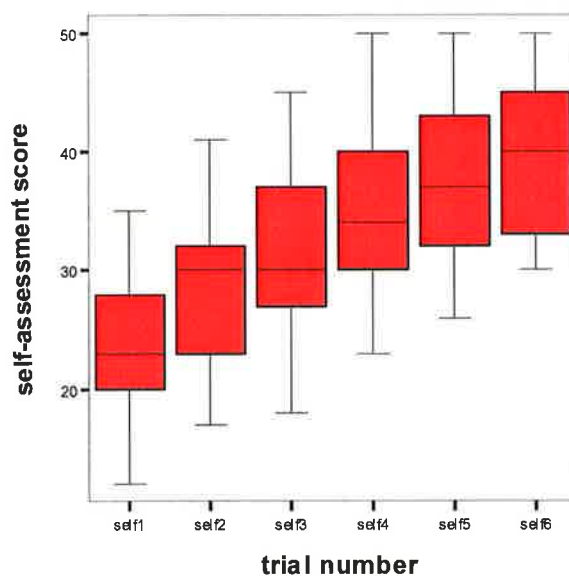


Figure 4.5 – self-assessment scores per procedure

When the scores for all 6 trials were summed for every candidate, the scores were compared by gender using Mann-Whitney-U tests, and male subjects scored themselves significantly higher than female (215.63 vs 178.56, $p = 0.046$)

Although the trend towards overall improvement over the course of the 6 trials matched the gradual improvement with self assessment scores, there was little

correlation when individual scores were compared to individual metrics. In all cases, the correlation was proportional....

The only scores to significantly correlate per trial were:

- Self-assessment with error score trial 4 (-0.542, p = 0.025)
- Self-assessment with error scores trial 5 (-0.486, p = 0.048)
- Self-assessment with time trial 4 (-0.523, p = 0.031)

4.3.7 Inter-metric correlation

Table 4.10 - Correlations between each of the 7 different simulator metrics used in the study

		time	contrast	fluoro	errors	placement accuracy	residual stenosis	% lesion cover
time	Correlation Coefficient	1.000	.299	.779(**)	.605(**)	.422	.444	-.410
	Sig. (2-tailed)	.	.200	.000	.005	.081	.149	.091
contrast	Correlation Coefficient	.299	1.000	.135	.358	.181	.077	-.170
	Sig. (2-tailed)	.200	.	.569	.121	.473	.811	.500
fluoro	Correlation Coefficient	.779(**)	.135	1.000	.643(**)	.370	.423	-.411
	Sig. (2-tailed)	.000	.569	.	.002	.130	.171	.090
errors	Correlation Coefficient	.605(**)	.358	.643(**)	1.000	.495(*)	.549	-.428
	Sig. (2-tailed)	.005	.121	.002	.	.037	.064	.077
placement accuracy	Correlation Coefficient	.422	.181	.370	.495(*)	1.000	.916(**)	-.895(**)
	Sig. (2-tailed)	.081	.473	.130	.037	.	.000	.000
Residual stenosis	Correlation Coefficient	.444	.077	.423	.549	.916(**)	1.000	-.979(**)
	Sig. (2-tailed)	.149	.811	.171	.064	.000	.	.000
% lesion cover	Correlation Coefficient	-.410	-.170	-.411	-.428	-.895(**)	-.979(**)	1.000
	Sig. (2-tailed)	.091	.500	.090	.077	.000	.000	.

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

We examined the internal consistency of the performance metrics measured by the simulator. Spearman's rho was used as it is a non-parametric correlation coefficient.

As can be seen from table 4.10, the metrics that significantly correlated included:

- Fluoroscopy and total procedural time
- Error scores and total procedural time
- Error scores and fluoroscopy time
- Error scores and placement accuracy
- Placement accuracy and residual stenosis
- Placement accuracy and percentage of lesion covered (negatively correlated)
- Residual stenosis and percentage of lesion covered (negatively correlated)

The high number of correlations supports the robustness of these metrics.

We also correlated the simulator error scores with the results from the video assessments and found several significant correlations.

Table 4.11 Correlations between simulator and video error scores

Procedure	Correlation coefficient	p-value
1	0.802	0.000
2	0.702	0.001
3	0.599	0.014
4	0.575	0.040
5	0.556	0.049
6	0.744	0.001

Table 4.12 Correlations between simulator error scores and video global assessment scores

Procedure	Correlation coefficient	p-value
1	-0.56	0.032
2	-0.769	0.001
3	-0.662	0.014
4	-0.576	0.040
5	-0.188	NS
6	0.471	NS

4.4 Discussion

The primary aim of this study was to assess the value of feedback, and to compare expert with non-expert feedback.

The results from chapter 2 show a clear benefit to feedback, but a particular kind of benefit - the more general skills associated with the procedure such as ease of instrument use and time were not reliant on feedback, but procedural specific errors were reduced. A similar pattern was seen here. All the subjects improved significantly as they performed the 6 procedures but for the general procedural scores, there was no difference in the rate of improvement in the subjects who did and did not receive feedback. However, there were significantly fewer errors committed by the subjects who received expert feedback, both when compared to Group 1, control, and Group B, who had feedback from a non-expert facilitator.

This demonstrates the extra benefit of expert feedback during simulator-based courses, both over no feedback, and feedback from a non-expert. It is true that the mere presence of an expert may also be a factor as the subjects in Group C may have felt under more pressure to perform due to the presence of an expert. Certainly some differences in performance were noted on trial 1, and subjects in group C tended to perform better. It was also clear during the data collection sessions that subjects in group C were very conscious of the presence of the consultant. However, the subjects in the other groups were also motivated to perform well and try to outperform their peers. Many subjects asked about the best score that had been achieved so far during the testing sessions and wanted to know how they ranked compared to the rest of the group. The generic instrument handling skills therefore seemed to be more independent of expert feedback. Hislop et al had similar findings, in that they separated time to complete a procedure (“innate endovascular skill”) from qualitative

assessment of procedure, which was based on observer ratings of performance, and found that previous experience and skill level correlated with better qualitative performance but not time to complete the procedure, suggesting the generic skills necessary to perform a such procedures quickly are separate from ability to perform the procedure well. Although there is some benefit in mastering the instrument handling skills in the simulator environment before proceeding to the real procedure – the “pre-trained” novice (Gallagher at al, 2005) - it is obviously optimal to also cultivate good, safe habits as there is a risk that the simulator could reinforce bad practice in the absence of feedback. (Gallagher et al, 2008) It could raise the confidence levels of novices while allowing them to practice serious errors. In addition, there is a certain amount of intrinsic feedback with regard to the generic performance measures in that the trainee knows how long they spend performing the procedure and how many mls of contrast they have used, therefore there may be little extra benefit from feedback regarding these performance measures.

All performances were recorded also to objectively assess performance. Some of the VIST metrics have been questioned regarding their validity and have not consistently been demonstrated to be construct valid. We assessed performance using a checklist for specific errors and a global assessment scale. Using the checklist there were significant differences between the 3 groups (figure 4.2), and differences with the global score also although these were less marked. There were clear patterns regarding error commission and most errors were due to inability to manoeuvre instruments rather than lack of cognitive knowledge. Interestingly, when the video-based assessment of error scores is examined, (arguably more clinically relevant than the simulator-based metrics), the differences between groups B and C was not significant (table 4.8), although both groups performed better than group A. This

suggests that while the provision of performance feedback appears to reduce errors, the status of who provides the feedback is not important as long as the facilitator is familiar with the VIST simulator. We did not see any significant reduction in errors when feedback was provided by an expert compared to a surgical research fellow with no previous vascular or endovascular experience. This is relevant for planning courses and teaching as the most common rate limiting factor ie availability of expert faculty may be less important for successful learning. As discussed above, there may also have been a Hawthorne effect in Groups C which is more difficult to measure. (McCarney R et al, 2007). As can be seen from figures 4.4a and b, there was also a clear improvement for these scores across the 6 trials. In this analysis, different errors were not weighted but it would be interesting if more clinically serious errors were weighted more heavily. As videos were recorded, these could also be a useful training resource for the trainee – they could watch taped performances on a separate occasion and score themselves, which would be a valuable learning resource.

As in chapter 2, the subjects were novice in the particular procedure and all showed a significant learning curve over the course of the 6 procedure repetitions. The improvements seen during a relatively short period of training in a group of novices is similar to the findings in chapter 3, and has been seen elsewhere in the literature (Aggarwahl et al, 2005, Neequaye et al, 2007). As a time limit of 40 minutes was used, the completion rate among the subjects for trial 1 was not 100%, in fact 7 of the 18 subjects failed to complete the procedure on their first attempt. However, all subjects were able to complete the procedure by their third attempt. This shows the benefit of repeated, structured practice, something which is difficult to achieve in the clinical environment. A trainee might typically have the opportunity to perform a particular procedure once during an operating list, and there may be a long time

interval before their next opportunity during which time their skills may have decayed. In addition, procedures vary hugely from case to case – simulator training gives the opportunity to practice an identical procedure several times, building confidence and skill.

While simulators can never replicate fully the clinical experience, the VIST simulator is an advanced, high fidelity model. The metrics are appropriate, and aim to reinforce good and safe habits. We demonstrated significant correlations between several of the different metrics. For example, balloon placement accuracy was highly correlated with residual stenosis at 0.961, meaning that the greater the distance from the centre of the balloon to the centre of the lesion, the greater the resulting residual stenosis. This is to be expected clinically and emphasises the content validity of the metrics used. Placement accuracy was inversely correlated with percentage of lesion covered for the same reason and interestingly was also correlated with error score meaning the less accurate the subject was with regard to balloon placement, the more likely they were to commit other errors. This also supports the robustness of the metrics used. Previous research has questioned the value of these metrics and some of the metrics could be criticised for their relevance. Interestingly, the video assessment scores correlated highly with the VIST-measured performance outcomes (figure 4.11 and 4.12) suggesting that if trainees were trained to access the performance report provided by the simulator after every performance they could potentially derive some of the same benefit from performance feedback provided by an observer.

Although none of the subject had performed any endovascular procedures before participating in the study, some subjects had rotated through vascular surgery jobs, and assisted at procedures. This had no effect on their performance in the study and did not confer any benefit. Similarly there was no obvious difference between subjects

who did or did not wear glasses or play video games. This may be because the four subjects who played video games did so infrequently, for less than 1.5 hrs week. In addition our numbers were small.

Subjects were asked to self-assess their own performance after every trial. Although individual correlations between self-assessment and particular metrics did not correlate significantly, it is clear from comparing figures 4.5 with figures 4.1a-d that the subjects' self-assessment mirrored their clinical performance. Self-assessment is an important part of the learning process, and helps trainees to improve performance on subsequent trials or practices. Self-assessment would also facilitate training in that it would avoid the need to have an expert, or indeed any assessor present throughout training. The literature however suggests that trainees tend to over estimate their abilities when self-assessing (Evans AW et al 2002, Evans AW et al 2007, Pandey VA et al 2008) It has also been suggested however that virtual reality training and objective feedback can improved trainees' ability to self-assess their performance. The self assessment by the subjects suggests accurate self assessment, and it was obvious during the data collection itself that trainees were aware when they had committed an error. As mentioned above, error commission tended to be due to manual difficulty handling the instruments rather than defective cognitive knowledge, and in some cases subjects were aware that they were committing an error repeatedly but unaware of the reason why and in these situations the feedback received was particularly beneficial. The improvement in self-assessment scores also suggests increasing confidence levels which may be beneficial to performance and the fact that the subjects were compelled to reflect on their performance was also likely to have been beneficial.

One of the important advantages to simulator based training is its efficiency and cost effectiveness, and removing some elements of skills training from the operating environment to the skills lab can save time and money as operating time in a fully equipped theatre is valuable. As there is a high cost and logistical difficulty associated with providing expert faculty for surgical training courses, it is advantageous to know if this is necessary for optimal learning to take place

4.5 Conclusions

In conclusion, we have demonstrated that short intensive training allows novices to make significant performance improvements. Generic skills are improved regardless of the availability, or provider of feedback meaning that even independent practice on a simulator is beneficial, but performance is freer from procedural errors when an expert provides feedback. Our video-based assessment suggest that non-expert facilitator can provide performance feedback as effectively as expert faculty meaning that training on the VIST simulator can be beneficial with the support of a technician only. Therefore while expert faculty are desirable at intense skills courses, such training has other benefits if faculty are not available. We have demonstrated the robustness of the simulator metrics, which are an important part of training and their incorporation into training sessions could maximise the benefit for the trainee. Objective assessment of performance correlated well with the simulator metrics, again supporting their use. Viewing of procedural videos could be incorporated into training

regimens. Trainee self-assessment reflected performance standard and suggests that mandatory self-assessment where the subjects are compelled to reflect on their performance may be beneficial.

Chapter 5

Training surgical skills using non-surgical tasks

– can Wii improve surgical performance?

Chapter 5 - Training surgical skills using non-surgical tasks – can Wii improve surgical performance?

5.1 Background

The procedural skills required to perform surgery today have changed enormously over the past number of years. Although certain specialities have been less affected, the advent of minimally invasive surgery and technological advances in the medical device industry have driven the change towards image-guided procedures. Much colorectal, upper GI and urological pathology is now treated with laparoscopic surgical techniques, and other specialities such as ENT and vascular surgery have seen similar changes.

The skills needed to perform image-guided surgery are not a natural extension of those required for open surgery and challenges included the fulcrum effect, reduced haptic feedback, impaired depth perception and visual cues, and the mental reconstruction of 3D from 2D imagery (Gallagher et al, 2003). The difficulties in acquiring these skills can lead to a longer learning curve for minimally invasive surgery(Tekis et al.,2003, Southern Surgeons Club 1991, Dunn et al 1994), which is further complicated by training hour restrictions (ACOG committee 2008, ACGME, Calman Report 1993, EWTD, www.dohs.ie).

It has been suggested that abilities in non-surgical tasks may translate to the surgical setting, with video-gaming attracting particular attention due to the obvious similarities in the manual skills required, such as hand-eye coordination and screen mediated task execution. It is possible that younger surgeons may acquire skills in laparoscopic surgery more rapidly than their elder colleagues, possibly because they have been exposed to video games at a young age(Tsai et al, 1994) Video games (VG)

have become extensively integrated into popular culture and anecdotal observations of young surgeons suggest that VG play contributes to performance abilities in laparoscopic surgery. Video games in themselves have become increasingly more advanced and more widespread. In early 2008 the NPD group reported that 72% of people age six to forty-four in the U.S. played video games in the year 2007 with most players using the computer to play video games. The average age of players across the globe is middle to late 20s, and is increasing as older players grow in numbers (Flew, 2005). The presence of a definite link between gaming ability and laparoscopic skills would have many training implications. It could effect selection for surgery, decrease training time and potentially be a cheaper and more accessible alternative to complex simulation labs.

Many recent studies have investigated the link between gaming and surgical skills, but most have investigated the influence of prior gaming habits on surgical ability. Some have shown conflicting or weak results. Rosser et al demonstrated that surgeons who had played video games in the past for more than three hours/week made 37% fewer errors, were 27% faster, and scored 42% better overall on a laparoscopic training course than surgeons who never played video games. Enocchson et al investigated the relationship between prior video game playing and performance on the GI Mentor simulator, and showed a positive effect on only 2 of the performance metrics, efficiency of screening and duration of procedure (Enochsson et al, 2004). However, Grantcharov et al (2003) showed that gamers made significantly less errors on a porcine laparoscopic cholecystectomy model than non-gamers, but that gender and hand dominance also affected the performance results and Madan et al (2008) showed that prior gaming was associated with better scores on a VR task only, not a box trainer task, but that the difference was not significant.

In a separate study, Madan et al (2005) showed that video games did not predict improved performance, and Glaser et al (2005) showed that any benefit associated with prior VG exposure was short-lived and disappeared with practice. Harper et al showed that prior VG experience was inversely correlated with ability to learn robotic suturing.

All of these studies have investigated the effect of prior gaming on surgical skill, and many have shown that conflicting factors such as musical or sporting ability may also have an effect.

While these results are interesting they have relevance only in selection and have less relevance for trainees or students who are already working in surgery as prior gaming experience cannot be altered. Fewer studies have prospectively exposed subjects to video games to measure the effect this may have on surgical ability. Rosenberg et al trained subjects for 2 weeks using an X-box video game and while there was a trend towards a superior performance than the control group, none of the performance metrics revealed a statistically significant advantage to VG training.

Kolga Schlickum et al trained subjects using a FPS 3 dimensional module and found that in comparison to subjects trained with a 2D game, the 3D game was associated with a better performance. However there was no control group for comparison, so this study was merely comparing the benefits of different kinds of video games. Given the challenges inherent in laparoscopic surgery, it seems intuitive that 3D games may be of more benefit than the more traditional video games.

There has been much recent media attention on the "Nintendo Wii" video game and suggestion that the unique bi manual, 3d nature of this game could improve surgical skills. The aim of this study was to investigate if a period of Wii practice would improve the performance of laparoscopic novices compared to a control group.

5.2 Materials and Methods

5.2.1 Setting

The study took place in the National Surgical Training Centre at the Royal College of Surgeons over a 6 week period in May and June 2009

5.2.2 Subjects

A group of 22 medical and science students were recruited to the study. Subjects volunteered following blanket emails to classes. There were no age limits but participants were ineligible if they:

- had any laparoscopic surgical experience (observing laparoscopic surgery was not regarded as an exclusion criterion)
- had ever played video games regularly
- had previously played a Nintendo Wii for more than one hour

5.2.3 Randomisation

After recruitment, participants were randomised to Group 1, control or Group 2, Wii. A simple block randomisation scheme was used with blocks of four in order to keep the numbers in each group as similar as possible.

5.2.4 Apparatus – Nintendo Wii

The Nintendo Wii is described in detail in chapter 2

5.2.5 Study Procedures

Overview

All participants attended a testing session at the NSTC (Session1) which lasted approximately one hour. After a gap of 5-7 days, participants attended for a second testing session during which all the tasks from Session 1 were repeated.

In between the 2 sessions, subjects in Group 2, (Wii) had 3 hours of structured practice on the Nintendo Wii.

5.2.5.1 Details of testing session

All participants who attended the testing session initially supplied some basic demographic data. They then performed the following battery of tasks which included two physical and one virtual task on the ProMIS surgical simulator (Haptica, Ireland), and aptitude tests.

Bead Transfer Task on the ProMIS surgical simulator.

This is similar to the basic peg transfer task on the fundamentals of laparoscopic surgery (FLS) programme. Trainees use real laparoscopic instruments in their left and right hands to transfer beads to different, numbered pots and to a small bag, similar to an endobag used during laparoscopic appendicectomy procedures. The task encompasses three levels. All instrument movements are assessed, and the simulator supplies results for time to complete the task, instrument path length (IPL) and instrument smoothness (IS). In addition, any beads dropped are recorded by an observer.

Fine Dissection Task on the ProMIS surgical simulator

This task requires the trainee to cut a standard triangular shape from a latex glove which is stretched over a frame within the simulator, using a laparoscopic cutter and grasper. In addition to cutting the shape as accurately as possible, trainees are instructed to avoid making any perforations in the bottom layer of the latex glove. As before, time, IPL, and IS are recorded by the simulator. In addition, the number of perforations made in the bottom layer of the glove was counted, and the quality of the dissected triangle was assessed in a blinded fashion and given a score between 0 and 5.

Instrument Handling task on the ProMIS surgical simulator – this purely virtual task requires the trainee to locate and touch various targets which appear on a virtual abdomen. After touching each target with the laparoscopic instruments, the trainee is required to hold the instrument in place for 5 seconds. As before, the metrics measured by the simulator for this task are time, IPL and IS.

Aptitude tests

All subjects underwent aptitude testing as described in chapter 2

5.2.5.2 Nintendo Wii practice

The Nintendo Wii was used as described above

The practice session was unsupervised but structured and based around 4 different games which are part of the “Super Monkey Balls” software package.

Four games in total were selected as part of the practice session for their ability to train different skills which are relevant to laparoscopic surgery.

“Whack a Mole”

This game was selected as it requires the player to use fine movements of the hand to the right and left, and also backwards and forwards in order to position the virtual tool over the target. Therefore depth perception is required, as is the ability to relate depth of field as perceived on a 2D monitor to guide hand movements.

During the practice session, subjects had to reach a minimum score, or practice for 15 minutes, whichever was achieved first.

“Alien Attacks”

This game was selected as it requires the player to use both hands in coordination, ie both the Wii Remote and Nunchuck are used. The hands have to move in similar motion to turn from left to right, and in opposite directions to rotate.

“Asteroid Crash”

This game was selected as it requires the player to hit certain targets with speed and accuracy. Targets flying towards the player must be destroyed before they reach the player, but there are “good” targets which must be avoided, requiring the player to aim with precision and speed. This game becomes progressively more difficult the longer the player continues.

“Dangerous route”

This game requires the player to manipulate a horizontal surface. By tilting the surface in various directions, the player must navigate a figure around a certain route. If the figure falls off the edge, the game must be started again at the beginning. The

tilting is performed in “real” space holding the Wii remote and the movements made by the player are transferred to the monitor.



Figure 5.1 – screen shot for “Whack- a Mole” game



Figure 5.2 – screen shot for “Alien Attacks” game

5.2.6 Statistical analysis

All data were entered into an SPSS database. Improvements between the two testing sessions were analysed using ANOVA for repeated measures, with group as a factor for comparison between the groups. Mean differences between the groups within one testing session were compared with Mann-Whitney-U tests. This test was used as the groups being compared were relatively small and thus it was felt that nonparametric analysis was more appropriate.

5.3 Results

22 subjects in total were recruited to the study; however 3 of these subjects dropped out during the course of the study and failed to complete the second practice session. Results are therefore calculated for 19 participants only.

Demographics are displayed in table 5.1.

The mean age of the subjects was 21.95 years. Two of the 19 subjects were left-handed.

No significant differences existed between the groups for any of the aptitude tests

All the subjects' performances improved from session 1 to session 2.

There were no significant differences between the control and the Wii group for any of the performance metrics measured in session 1.

Table 5.1 - demographics

Mean age	21.95 (sd 3.08)
Mean year medical school	2.67
Handedness	17 right, 2 left

5.3.1 Results for Bead Transfer Task

All subjects improved from session one to session 2. (Figures 5.3a-c) Time taken was shorter, and IPL and IS were significantly reduced.

Figure 5.3a – overall results for time for the Bead Transfer Task

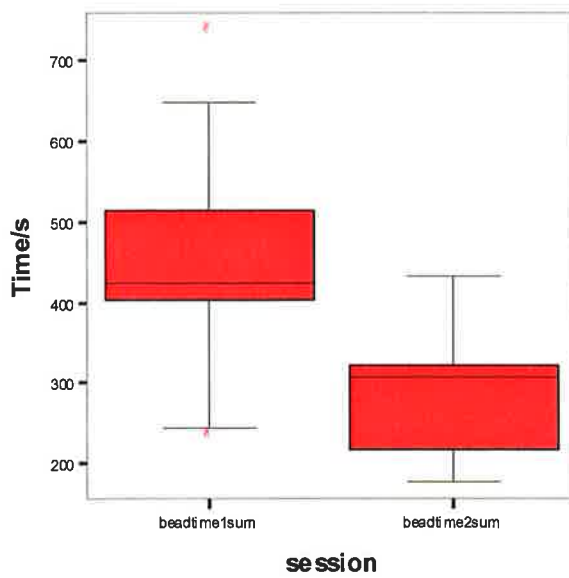


Figure 5.3b – overall results for IPL for the Bead Transfer Task

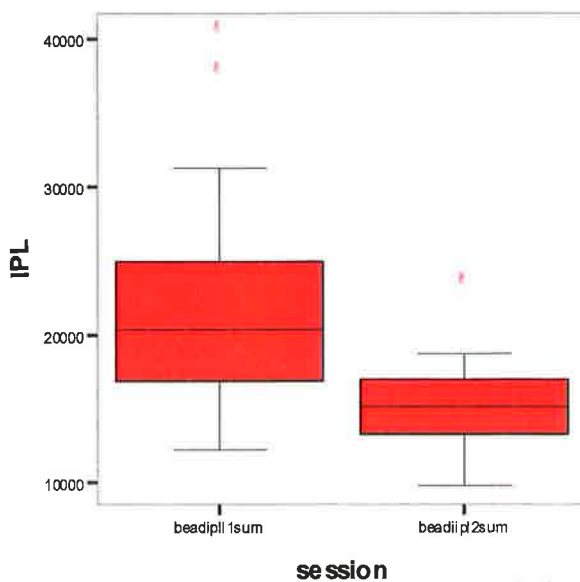
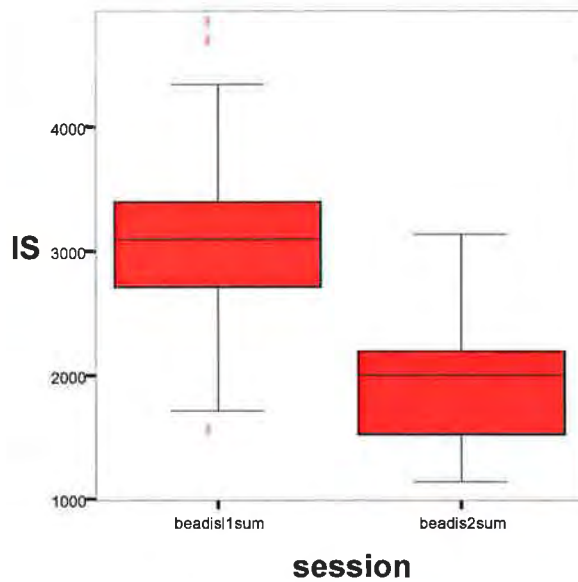


Figure 5.3c – results for IS for the Bead Transfer Task



When the control group was compared to the Wii group, there was no significant difference in the improvement in instrument scores seen from session 1 to session 2 (using ANOVA for repeated measures). However, for bead dropping, the Wii group improved significantly from session 1 to session 2 when compared to the control group ($p = 0.045$)

The Wii group on average had better instrument movement scores in session 2, although the difference was not significant (Table 5.3)

Table 5.2 - improvements in performance metrics from session 1 to session 2, bead transfer task

	ANOVA – within subject improvement	ANOVA - within subject improvement*group
time	$p = 0.000$	$p = 0.418$
IPL	$p = 0.000$	$p = 0.269$
IS	$p = 0.000$	$p = 0.278$
Beads dropped	$p = 0.193$	$p = 0.045$

Table 5.3 – actual instrument scores, bead transfer task

	Control group	Wii group	p- value (Mann – Whitney-U)
IPL – session 2	1578	1409	0.730
IS – session 2	1996	1962	0.666
Time – session 2	298	296	0.796
Beads dropped	1.0	0.66	1.00

5.3.2 Results for Glove Cutting task

A similar pattern was seen for the glove cutting task – when all 19 subjects were analysed together, there was a significant improvement on all metrics from session 1 to session 2, but when the two groups were compared, there was no significant differences on the amount of improvement seen (using ANOVA for repeated measures, table 5.4), although the results for instrument smoothness approached statistical significance. However, the trend was towards a superior performance in the Wii group, as their mean performance in session 2 was better than the control group, although not significantly so.

Table 5.4 - improvements in performance metrics for glove cutting task, session 1 – session 2

	ANOVA – within subject improvement	ANOVA within subject improvement*group
time	p = 0.005	p = 0.470
IPL	p = 0.013	p = 0.852
IS	p = 0.013	p = 0.067
No. of perforations	p = 0.035	p = 0.811
Quality of triangle	p = 0.009	p = 0.564

Table 5.5 – mean performance results for glove cutting task, session 2

	Control	Wii	Mann-Whitney-U
time	435	406	p = NS
IPL	13200	11698	p = NS
IS	3198	2793	p = NS
No. of perforations	1.89	1.0	p = NS
Quality of triangle	3.3	3.6	p = NS

5.3.3 Results for ProMIS virtual task.

The pattern of results for the virtual task was slightly different. Similar metrics were used to measure performance, ie time, IPL and IS, but there was no significant improvement seen from session 1 to session 2 for all the subjects taken together, and also no significant improvement when the subjects were compared by group. There were no significant differences between performance results in session 2, and the Wii group performed better for time and IPL only

Table 5.6 Mean performance results for virtual task, session 2

metric	Control	Wii	p-value (MWU)
Time/s	385	371	NS
IPL/mm	7,300	7270	NS
IS	1075	1135	NS

5.4 Discussion

There remains little debate regarding the value of surgical simulation in the training of surgical skills. However simulators are expensive and therefore scarce. In Ireland for example, although some hospitals have purchased simulators, there is only one general surgical skills lab in the country, and one mobile skills unit. This has to cater for trainees who may be in peripheral units. This level of exposure to skills training is clearly inadequate. The ability to effectively train surgical skills using non-surgical tasks would therefore be extremely beneficial. Video games in particular are increasingly cheap and available and the literature suggests that previous exposure to video games, along with other physical activities may be beneficial to surgical skills. Despite the recent interest in the effect of non-surgical skills on surgical performance, there is little data which prospectively investigates the effect of VG playing on surgical skills. Most studies correlate self-reported previous VG playing with surgical performance with mixed results. However, like gender and handedness, there is little practical benefit from these observations (other than selection into surgical training) if it cannot be demonstrated that prospective VG training improves surgical technical skills.

The purely technical aspects of laparoscopic surgery would appear to have an overlapping skill set with those required to play 3D video games. The recent advances in gaming technology have made such devices more widespread, and there have been suggestions that games requiring 3d movements may have more relevance to surgeons than more traditional 2d video games (Rosenberg et al, SARS ref). While video gaming may have the potential to positively affect surgical skills, they have the potential to impact solely on the technical aspects of surgery, which are only a small part of the many skills required to operate. However, if the trainee can approach laparoscopic surgery with some of the ergonomic challenges resolved they will be

able to devote far more attention to learning about the other, non-technical aspects of operating.

In our study, although the trend was towards a better post training performance in the Wii group (who performed better in session 2 for 11 of the total of 12 performance metrics), the only result to reach statistical significance was the score for improvement in dropped beads in the Pro MIS Bead Transfer Task. It is of interest however that there was a large and significant improvement in the complete cohort from session 1 to session 2. It is possible that the practice schedule was insufficient – subjects in the Wii group had only 3 hours practice which in some cases was dispersed over 7 days between the 2 sessions. A more regular practice schedule such as 1 hour a day for 2 weeks may have yielded more significant results. In addition, if the time gap between the sessions had been more standardised better results may have been obtained. Scheduling difficulties meant that the time gap between the 2 sessions ranged from 3 to 7 days and although this did not differ significantly between the Wii and control group, this may have affected the results. A last factor may have been the novice status of the subjects. As none of the subjects had even handled laparoscopic instruments before their first session, the learning curve was steep between session 1 and 2 (figures 5.3 a-c). The improvement in session 2 for all subjects was therefore large and may have disguised a smaller beneficial effect of the Wii training. Allowing for these limitations, the improved performance in the Wii group (for 11 of the 12 metrics) after only 3 hours of practice indicates that there may be an advantage to video game practice. Given the trend towards improved performance in the Wii group, it would be of benefit to investigate this effect further, possibly with trainees of different standards and with a heavier and more regular training programme in the Wii group.

The Wii itself was portable, easy to set up and easy to use. It did not necessitate the use of a specialised skills lab and could be used in any area with approx 1.5 m² of space. There are no consumables required to use it. Performance results can be saved and recorded manually, although they consist purely of outcome measures and not qualitative metrics. All subjects reacted positively to using the Wii during the course of the study. As video game playing is a recreational activity, trainees may be more inclined to practice and more importantly, any time spent practising on the Wii would not take away from already reduced working/training hours.

5.5 Conclusions

In conclusion, laparoscopic novices who were prospectively randomised to 3 hours practice on the Nintendo Wii demonstrated an improved performance on basic laparoscopic tasks when compared to controls, although the difference was only significant for the improvement measure on one performance metric on the Bead Transfer Task. This suggests that video gaming may positively impact on laparoscopic surgical skills and has implications for surgical training.

Chapter 6

Retention of surgical skills

Chapter 6 Retention of surgical skills

6.1 Background

A criticism of many short term, intensive surgical skills courses is that although short-term improvements in performance are often seen, the skills gained may not be retained in the long-term.

Retention of motor skills has been extensively studied and many variables are known to affect the retention of learned motor behaviours over lengthy no-practice intervals.

In one review, the variables that may affect the retention of motor skills were dichotomized into task variables and procedural variables. The task variables that may underlie the long-term retention of motor skill include (a) duration of the no-practice period, or retention interval; (b) nature of the response required to accomplish a particular motor task; (c) degree to which the learner can organise or impose order upon the elements that define the task; (d) structure of the training environment; and (e) initial or 'natural' ability of the learner in performance of a task without prior practice. The procedural variables that may affect the long-term retention of motor skill include (a) degree of proficiency attained by the learner during initial training; (b) amount and kind of refresher training; (c) transfer of skills on one task to performance on another task; (d) presence of interfering activities; (e) distribution of practice during training; (f) use of part-task versus whole-task training methods; and (g) introduction of extra test trials prior to final testing.

Many of these variables are incorporated into surgical skills courses. In The HALC training for example, the procedure was broken down into discrete steps, the training environment had no distracting factors, the subjects were novice in the procedure but had prior ability in surgery and they reached proficiency in the error scores within the initial training session.

One factor which we could not control and which is known to adversely effect motor skills retention was the massed nature of the training. It has been shown in studies that amount of practice per se does not affect learning and retention of the task. Rather, distribution of practice over several days was the most important factor affecting learning and retention. (Savion LH et al, 2005) It is hypothesised that passage of time is essential for a maximum benefit of practice to be gained, as the time delay may allow for consolidation of learning, possibly reflecting plastic changes in motor cortical representations of the skill.

Behavioural psychology (e.g., Guthrie, Hull, Skinner) emphasized practice variables in sensory-motor skills such as massed versus spaced practice, part versus whole task learning, and feedback/reinforcement schedules. Long-term retention of motor skills depends upon regular practice; however, continuous responses show less forgetting in the absence of practice than discrete or procedural skills. Repetition after task proficiency is achieved (overtraining) and refresher training reduces the effects of forgetting. Unlike verbal learning, sensory-motor learning appears to be the same under massed and spaced practice. Learning and retention of sensory-motor skills is improved by both the quantity and quality of feedback (knowledge of results) during Retention of skills has been extensively studied with the surgical literature, The relevance to surgical practice is obvious – trainee surgeons must be able to perform operations spaced through time to a continuously high standard and not merely as an isolated event. In addition, given the pressures of work-hours and training, surgical trainees are frequently called upon to perform procedures and operations that they may not have been practising regularly, this is often the case with emergency cases which are frequently most challenging as the patient may be very unwell or present with extreme disease.

Intense training courses are logistically easier to provide than courses in which practice is more distributed, however despite this it is known that distributed practice does improve retention (Moulton et al, 2006). Some studies have found evidence of retention even when skills were acquired during a once-off, intensive training course. In one study where novices were trained on an endoscopic sinus surgery simulator, skills were retained at 2 months (Uribe et al, 2004).

There is some evidence showing that if trainees reach proficiency as part of the training they are more likely to retain skills. Subjects who perform a set amount of tasks or trials without having to reach a certain standard may not retain skills even when the training sessions are distributed. (Howells et al, 2009, Maharaj et al, 2007)

Stefanidis et al showed that novices trained to proficiency on a simulator showed significantly better skill retention at 5 months when compared to controls, and importantly showed a better performance in the real operating room, although there was a slight deterioration in performance when compared to the last performance.

(Stefanidis et al, 2005) Novices also reached proficiency in the study by Uribe et al.

Other research has shown a partial retention following proficiency based training. In one study, when retested at 6 months, trainees showed retention for some tasks only and tended to do worse for tasks requiring fine motor skills than for more general tasks. However as a group, significantly less subjects reached the level at retention testing that they had achieved after training to proficiency (Sinha et al, 2008). In

another study, residents trained to proficiency required some retraining when assessed at 6 months, but displayed partial retention, and required even less retraining when tested at 12 months (Castellvi et al, 2009). Therefore, while retention of skills is not complete, partial retention makes it easy to retrain.

6.2 Methods

Retention testing was performed in a selection of the subjects who participated in the HALC and endovascular training.

6.2.1 Endovascular subjects

5 subjects in the endovascular study returned for retention testing. They were all subjects from group A. All of the subjects repeated one trial on the endovascular simulator, as described in chapter 4. None of the subjects had performed any endovascular procedures between the training and retention sessions.

6.2.2 HALC subjects

We tested the retention of 8 of the subjects who participated in the HALC training. They were all subjects in the feedback group. They completed 5 HALC cases as described in chapter 3, with standardised feedback after every trial.

They were retested after a period of between 2 and 4 months on 2 further simulated HALC procedures and their performance was assessed as described in Chapter 3. We also collected data on their operative activity between the 2 testing sessions.

6.3 Results

6.3.1 Results for endovascular subjects

7 performance metrics were analysed as described in chapter 4 – procedure time, volume of contrast use, fluoroscopic time, error score, placement accuracy of the balloon, residual stenosis and percentage of lesion coverage. The mean time between

the initial training session and the retention session was 15.8 weeks (table 6.1). None of the subjects performed any endovascular procedures (either real or simulated) between the training and retention sessions. Partial retention of skills was observed – although the mean performance at retention was poorer than the performance on trial 6, it was better than the mean performance at trial 1. Only 2 of the 5 subjects had completed the procedure within the time limit of 40 mins on trial 1, but at retention testing all subjects completed the trial in less than 40 mins.

Table 6.1 – subject details

subject	Weeks between training and retention	Endovascular procedures between training and retention	Trial 1 completed (in 40 mins)	Retention trial time/mins
1	18	0	no	32
2	15	0	yes	22
3	13	0	no	22
4	17	0	no	32
5	16	0	yes	27

Table 3.2 shows the mean scores obtained at trial 6, retention testing, and the trial number at which these scores were achieved during the initial training. Retention was complete for the error scores, ie subjects achieved the almost the same scores as they did for trial 6 at their initial training. There was a high score at trial 6 for placement accuracy due to an outlier, so the final score for this metric is worse than the retention score.

Table 6.2 – mean scores at trial 6, retention testing, and the trial number at initial training at which the retention score was achieved

metric	Mean score trial 6	Mean retention score	Trial at which this score was achieved
Time/s	1058.4	1668.2	Btw T4-5
Contrast use/ml	24.9	31.6	Btw T4 - 5
Fluoro use/s	645	915.4	Btw T4 - 5
Placement accuracy/mm	5.8	3.6	T5
Residual stenosis	19.5	25.6	Btw T 3-4
% lesion covered	88	85	Btw T 3-4
Simulator error score	16.5	16.4	T6

Figures 6.1 a-c show the mean results for fluoroscopy and contrast use and error score for each of the 6 trials and the retention trial. Table 3.3 shows the differences between the performance results at trial 1 and retention when compared with Wilcoxon Ranks tests. The retention performance was significantly better than trial 1 for time, fluoroscopic time and error scores.

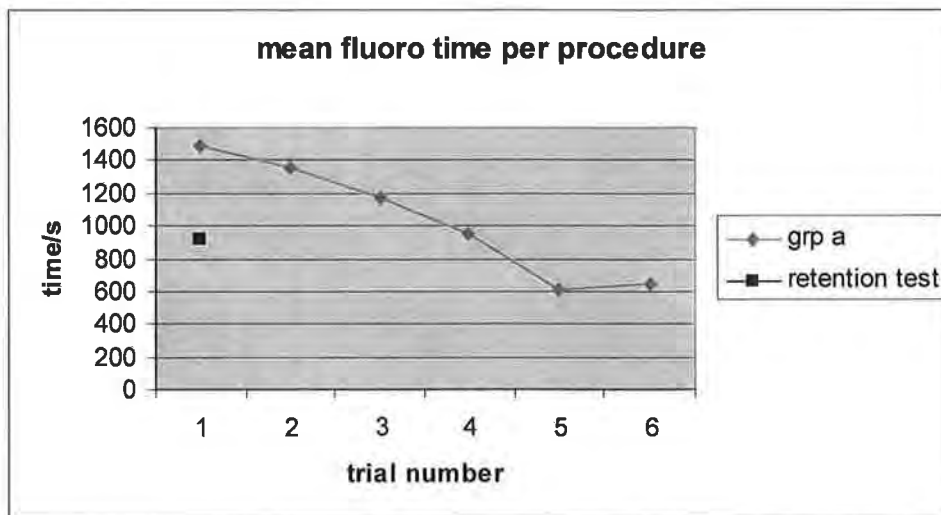


Figure 6.1a – mean fluoro time per procedure for trials 1-6 and retention

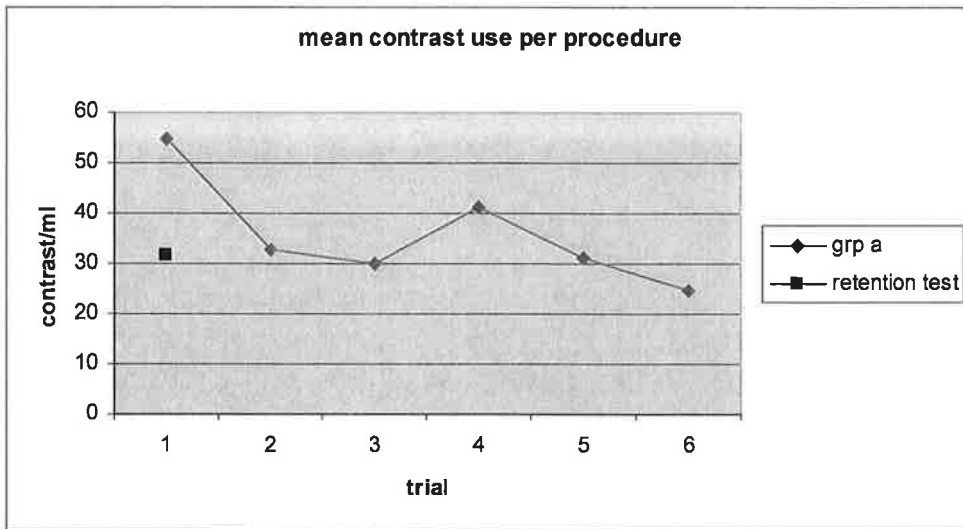


Figure 6.1b – mean contrast volume per procedure for trials 1-6 and retention

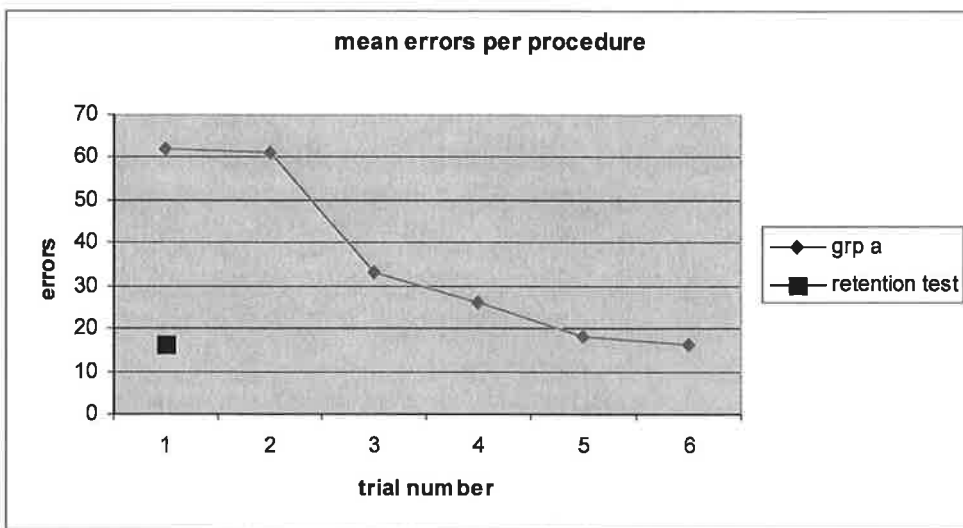


Figure 6.1 c – mean error scores per procedure for trials 1-6 and retention

Table 6.3 Improvement between trial 1 and retention – Wilcoxon Rank test

metric	Improvement between trial 1 and retention : p-values
Procedural time	0.028
Contrast volume	NS
Fluoroscopic time	0.028
Error score	0.028
Placement accuracy of balloon	NS
Residual stenosis	NS
Percentage of lesion covered	NS

6.3.2 Results for HALC subjects

8 subjects returned for retention testing.

Table 6.4 shows details of the time interval between testing sessions and their operative activity.

Table 6.4 – details of procedures performed

Subject code	Time/wks	Procedures performed between training and retention							
		Lap app	Lap chole	other lap	Non-adv lap	Adv lap	Open colectomy	endoscopy	other
1	25	40	20	1	61	2	3	300	11
2	25	0	0	0	0		0	1	33
3	17	12	7		19		2		
4	13	5	7	2	14	1	0	216	
5	10	13	7	0	20		9	80	5
6	9	16	9	5	25		0	40	33
7	8	4	2	0	10		2	14	
8	9	2	2	0	4		4	0	12
mean	14.5	11.5	6.75						

The mean time interval between the 2 sessions was 14.5 weeks. (Range 8-25 weeks)

In the intervening period, no subject performed any HALC procedures and only 2 subjects performed any advanced laparoscopic procedures. No subjects had any experience performing the simulated HALC procedure between the 2 sessions.

Figure 6.2a – line graph showing average IPL results for group 1, group 2 and retention testing

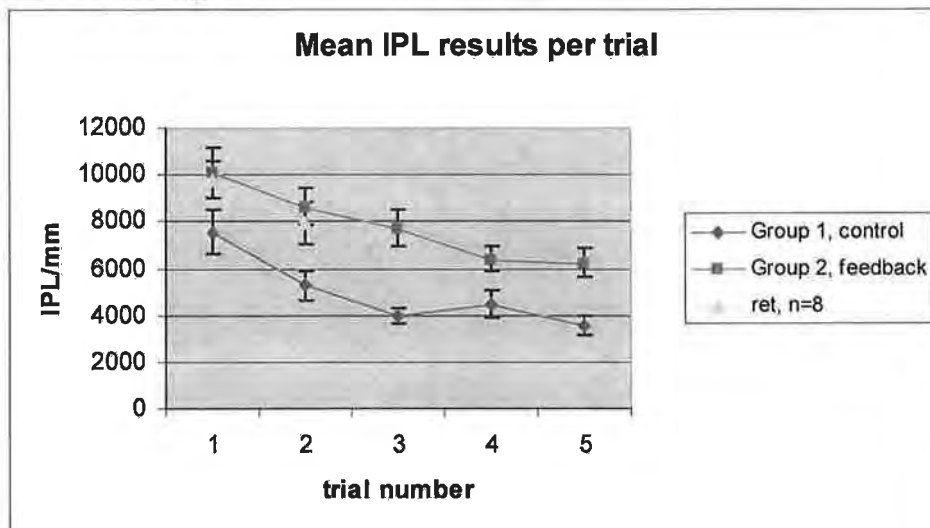


Figure 6.2b – line graph showing average IS results for group 1, group 2 and retention testing

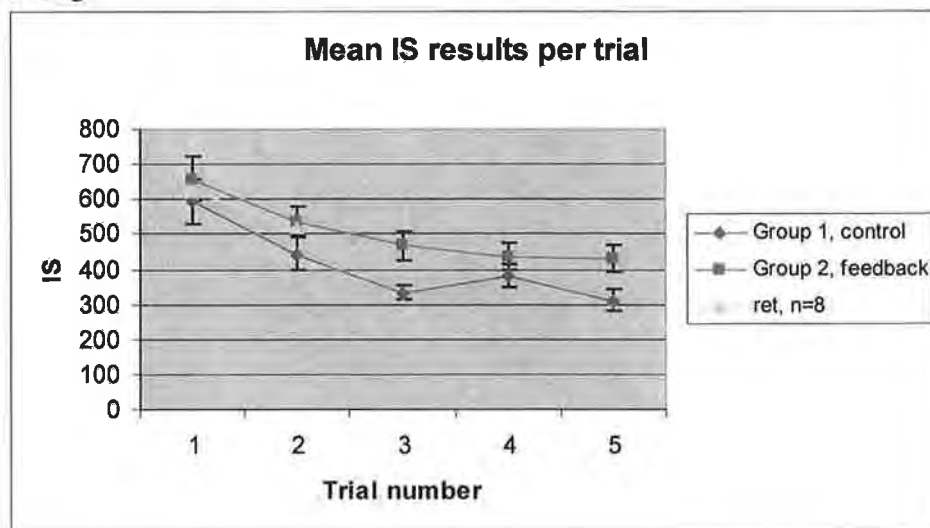
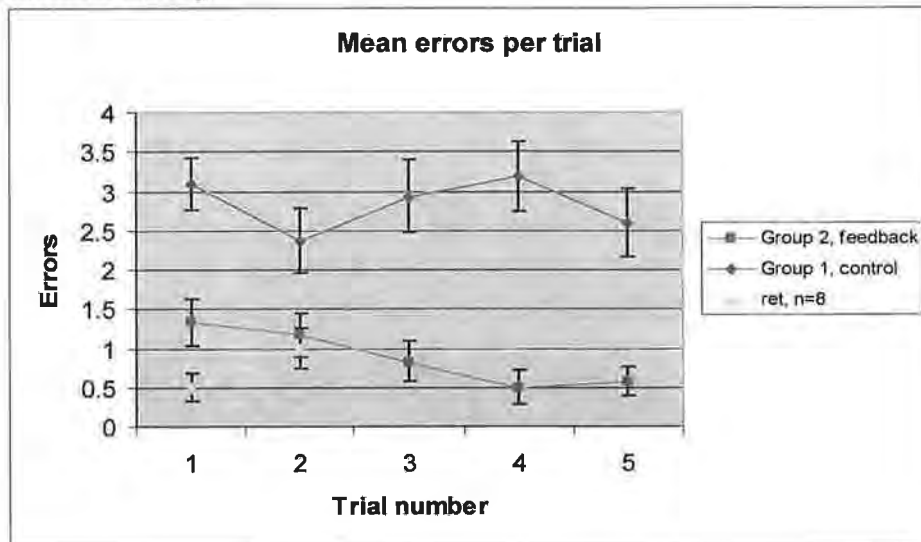


Figure 6.3c – line graph showing average error results for group 1, group 2 and retention testing



The results for the retention testing are displayed in figures 6.2 a-c. As can be seen from the graphs, a different picture was observed for the instrument movement results and the error scores. Although subjects performed their retention testing better than their initial trial, their performance was disimproved compared to their 5th trial.

Table 6.3 shows individual subject results for instrument path length retention. The first column shows the percentage difference between each subject's score on their 5th trial and their average performance over the 2 retention trials. Subject 1, for example scored a path length that was 10% longer (worse) than the path length score he achieved for trial 5. Subject 6 showed the largest disimprovement. Column 2 shows the difference between each subject's IPL result for their first trial and their retention score. Every subject bar subject 1 showed an improvement, indicating that some skill was retained and that the subjects were not starting at "square 1"

Table 6.5 – percentage improvements per subject in IPL between trial 5 and retention testing, and trial 1 and retention testing

Subject	IPL5+ret	IPL1+ret
1	-10.21	42.2
2	-35.51	15.41
3	-36.33	24.84
4	-57.15	4.65
5	-35.15	42.82
6	-94.31	14.69
7	-67.8	-39.74
8	-50.16	14.18

When the results were compared using Wilcoxon Ranks for repeated measures, the subjects improved their IPL and IS scores significantly from their 1st to 5th trials ($p = 0.012$, $p = 0.012$) but there was no significant improvement between their 1st or 5th trial, and their first retention trial.

They improved their IPL scores significantly between their 1st and 2nd retention trials ($p = 0.093$), but not between their 1st and 2nd original trials

There was stronger evidence of retention for the error scores. The mean error score for the 5th trial was 0.71 and for the 1st retention trial it was 0.5. There was a significant difference between the error score on the subjects' 1st trial and 1st retention trial.

Using a linear regression model, neither elapsed time, nor procedures performed between the initial testing session or retention session were significantly associated with performance on the first retention trial.

When time elapsed and procedures performed were correlated with performance on the first retention trial there was no clear relationship seen. For example, the number of procedures performed was negatively correlated with path length on the first retention trial expected, but not significantly so, and was significantly positively correlated with instrument smoothness, ie the more operative activity between the 2 sessions, the higher (worse) score obtained for instrument smoothness.

6.4 Discussion

Similar patterns were observed for skill retention in both data sets, ie the subjects performing the endovascular and HALC procedures. In both groups, the initial training session took place within a short time frame and represented massed practice with the majority of subjects completing all trials within one day. All subjects then returned for retention testing with no simulator training between the 2 sessions, and no real experience of the procedure between the 2 sessions, although some of the subjects in the HALC group performed advanced laparoscopic procedures which involve an overlapping skills set. In both groups skill retention was partial, and was better for specific error scores than for the more general measures such as efficiency. However, the skill retention in the endovascular group was more impressive. They performed at a level which was the level achieved at about trial 4 during their intensive training session (table 6.2). The retention of error scores was particularly marked in the endovascular group and occurred despite the absence of performance feedback. On average they performed at the same level for error commission as on trial 6, despite no practice in between the sessions. All subjects who returned were in group a, which was the control group. However some of the errors committed were obvious to the trainees at the time as they had been discussed in the didactic teaching

session which all subjects received and the improvement in error scores may represent improvements in instrument handling skills which helped the trainee to avoid committing the errors, rather than a cognitive awareness of the errors committed. The mean time interval before retention testing in the endovascular group was 15.8 weeks, whereas many of the studies examining retention testing in the literature have a time interval of 6 months. Had our time interval been longer it is possible that the retention would have been lessened.

On examination of the HALC data, retention appears poorer when figures 6.1 and 6.2 are compared. However, these are different metrics and efficiency of instrument use cannot be directly compared to endovascular measures such as volume of contrast fluid used.

Given that none of the subjects performed any HALC procedures in the interval between the 2 sessions, the relative lack of retention is unsurprising. However, absolute skill decay did not occur; with the exception of 1 subject, all of the subjects performed better on their first retention trial than they did on their first trial at the initial testing session. If we had required the subjects to perform more than 2 trials at the retention session we could speculate that they would in fact have shown an improved learning curve. This pattern has been shown in the literature, ie that subjects partially retain skills and take less time to retrain (Howells et al, 2009, Stefanidis et al, 2005 and Sinha et al, 2008)

The fact that the retention for the error scores showed a different pattern is likely because the subjects attained proficiency for the error scores, and therefore the skills learnt were more resistant to decay. It may also be due to the structured feedback they received which focussed primarily on error commission. It is known that self-assessment and feedback enhances retention of skills and this is another advantage to

designing courses with these elements in place – not only will learning take place in a shorter time but the skills learnt are more likely to be retained in the presence of feedback.

Interestingly there was no correlation in the HALC subjects between number of operative procedures performed before retention and degree of retention. This may be because none of the subjects performed a HALC procedure between the two training sessions. However, most of the subjects performed laparoscopic procedures, which may have been expected to improve performance on the HALC as many of the motor skills required to perform hand-assisted laparoscopic surgery overlap with pure laparoscopic surgery.

Another point of interest in the HALC group is the huge variety in the volume and type of the procedures performed during the interval before retention testing. This highlights two of the problems with current training programmes. Firstly training is very unstandardised, with much depending on the trainee's site and trainer. The other is that because higher surgical trainees tend to rotate through different areas for one year, they may lose skills acquired in a previous rotation as there may be little opportunity to perform procedures outside the area of the speciality they are rotating through. This highlights the importance of skills lab training which can bring some standardisation and continuous reinforcement to training, regardless of the speciality in which the trainee is currently working.

6.5 Conclusions

In conclusion we demonstrated partial skills retention at intervals of 14 and 15 weeks despite initial massed training and no reinforcement of the skills between training and retention. Subjects' retention performance was inferior to their post training

performance for general metrics but superior to their first attempt. Attaining proficiency in some metrics led to 100% skills retention, and error scores showed less decay for both procedures. Feedback and competency-based training therefore appears to enhance retention in addition to shortening the initial learning curve. Despite time constraints which favour massed over distributed training, skills lab training trains skills which are resistant to decay.

Chapter 7

The benefits of a proficiency-based simulator training curriculum for colonoscopy

Chapter 7 The benefits of a proficiency-based simulator training curriculum for colonoscopy

7.1 Background

14.2 million colonoscopy and 2.8 million flexible sigmoidoscopy procedures were performed at 8207 medical centres in the USA in the year 2002 (BSG Working Party Report, 2001). This is already a huge burden on the health system and given the current recommendations for screening the amount of colonoscopies carried out and looks set to increase further.

Despite the frequency with which colonoscopies are performed, there are huge variations in performance standards. This is highlighted by reports describing variations in parameters such as caecal intubation rates and polyp detection rates (Cotton PB et al, 2003) These variations in technical skill may contribute towards higher cancer miss rates. In addition, although serious complications such as perforation are rare (incidences from 0.07 to 0.12% have been reported (Iqbal CW et al 2008, Luning TH et al 2007)), and the majority of the iatrogenic perforations occur following interventional rather than diagnostic procedures, the morbidity is high, and hence all efforts should be directed toward addressing potential causes for iatrogenic perforation including those related to the provider's skills.

Apprenticeship training is a lengthy process and increasingly unworkable in the current setting with increasing concern surrounding issues such as patient safety and

economic viability. With the current limitation in time, space, and availability of mentors, feasible alternative to the current apprenticeship training model are required especially early in the learning phase of novice, as this is when errors are most likely to occur.

One of the alternatives to the traditional teaching model is the use of proficiency-based, comprehensive curricula which incorporate training on virtual reality (VR) simulators. The simulation curriculum will be delivered early in the training programmes and will complement the currently used training model. These curricula have the potential to transform the training of future colonoscopy providers.

Research in other areas has shown that trainees who train on proficiency-based progression programmes which incorporate the use of VR simulation will perform better in the real operating room when compared to trainees who have undergone traditional training (Seymour N et al, 2002, Ahlberg G et al, 2007)

In the last 10 years several studies have investigated the effect of the use of virtual reality simulator training for colonoscopy on the trainees' skills acquisition and transfer to real cases.

In 1998 Tuggy tested 10 residents on a virtual reality flexible sigmoidoscopy simulator and showed that his simulator trained group performed better on their 1st colonoscopy. In 2002 Sedlak et al at the Mayo Clinic studied the effect of using a computer based simulation curriculum on the training of novice colonoscopists. The study showed that 25 simulated colonoscopy procedures over 10 hours were needed

for the novices to reach the set proficiency level. Although the number of subjects tested was small the simulation curriculum used was more comprehensive compared to those used in similar studies. Ferlitsch et al studied the effect of an intensive 3-week virtual reality endoscopic simulation training on novices. The GI Mentor simulator used successfully managed to differentiate experts from novices and helped the novices acquire the necessary skills for safe colonoscopy in a short time (Ferlitsch et al, 2002)

In 2007 Park et al presented a well constructed randomised controlled trial which demonstrated the successful transfer of virtual reality simulator acquired colonoscopy skills to colonoscopy procedures performed on patients. Although the study was sufficiently powered to show a significant difference between trainees trained on a simulator, and traditionally, the simulation curriculum used was insufficient, and subjects trained on the simulator-based curriculum were not required to reach proficiency before they commenced their live cases. In addition, subjects were tested only on their first colonoscopy. Although it is intuitive that the differences between the two groups will start to decrease after the first case as those in the traditionally trained group will learn as they continue to perform the procedure, it is difficult to know how many cases this will take. Cohen et al have suggested the benefit may persist for as many as 80 cases.

7.2 Materials and methods

This was a prospective, randomised, controlled, blinded study.

Ethical Approval: Independent Ethics Committees Approval was obtained from the Royal College of Surgeons in Ireland and from each nominated study centre prior to commencement of the study.

7.2.1 Aims

Primary objective: To investigate if trainee colonoscopists trained with a proficiency-based curriculum which incorporates VR simulation, will perform better over their first five colonoscopies than those who only received standard or traditional training.

Secondary objective: To design and deliver a dedicated curriculum for colonoscopy training

7.2.2 Setting

The National Surgical Training Centre at RCSI

The following university teaching hospitals:

Beaumont Hospital, Beaumont

St Michael's Hospital, Dun Laoghaire

Adelaide and Meath Hospital, Tallaght

James Connolly Memorial Hospital, Blanchardstown

7.2.3 Simulator

The simulator used in the study was the GI Mentor II by Simbionix, Cleveland, USA, as described in Chapter 2

7.2.4 Participants

There were 3 groups of participants in the study – subjects, patients and consultant evaluators

7.2.4.1 Subjects

The subjects of the study were medical and surgical trainees.

Inclusion criteria

- Surgical and medical trainees working in the nominated study centres.
- Post-graduate year (PGY) 2-4

Exclusion criteria:

- Performed any complete colonoscopy procedures as the primary endoscopist.
Any trainee who had reached the caecum independently before the study was excluded. This did not include previously performing part of a colonoscopy or observation of any live cases.
- Performed >10 upper endoscopy procedures
- Performed any advanced endoscopy procedures.
- Previous proficiency based training on an endoscopy simulator. (A chance to try a simulator in the context of a one-day colonoscopy course was not regarded as an exclusion criterion as the exposure is typically minimal.)

7.2.4.2 Patients: Patients scheduled for elective colonoscopy were screened for suitability to participate in the study and were given an information leaflet and a consent form to sign. (appendix 7.3)

Inclusion criteria

Patients scheduled for full, elective colonoscopy.

Exclusion criteria

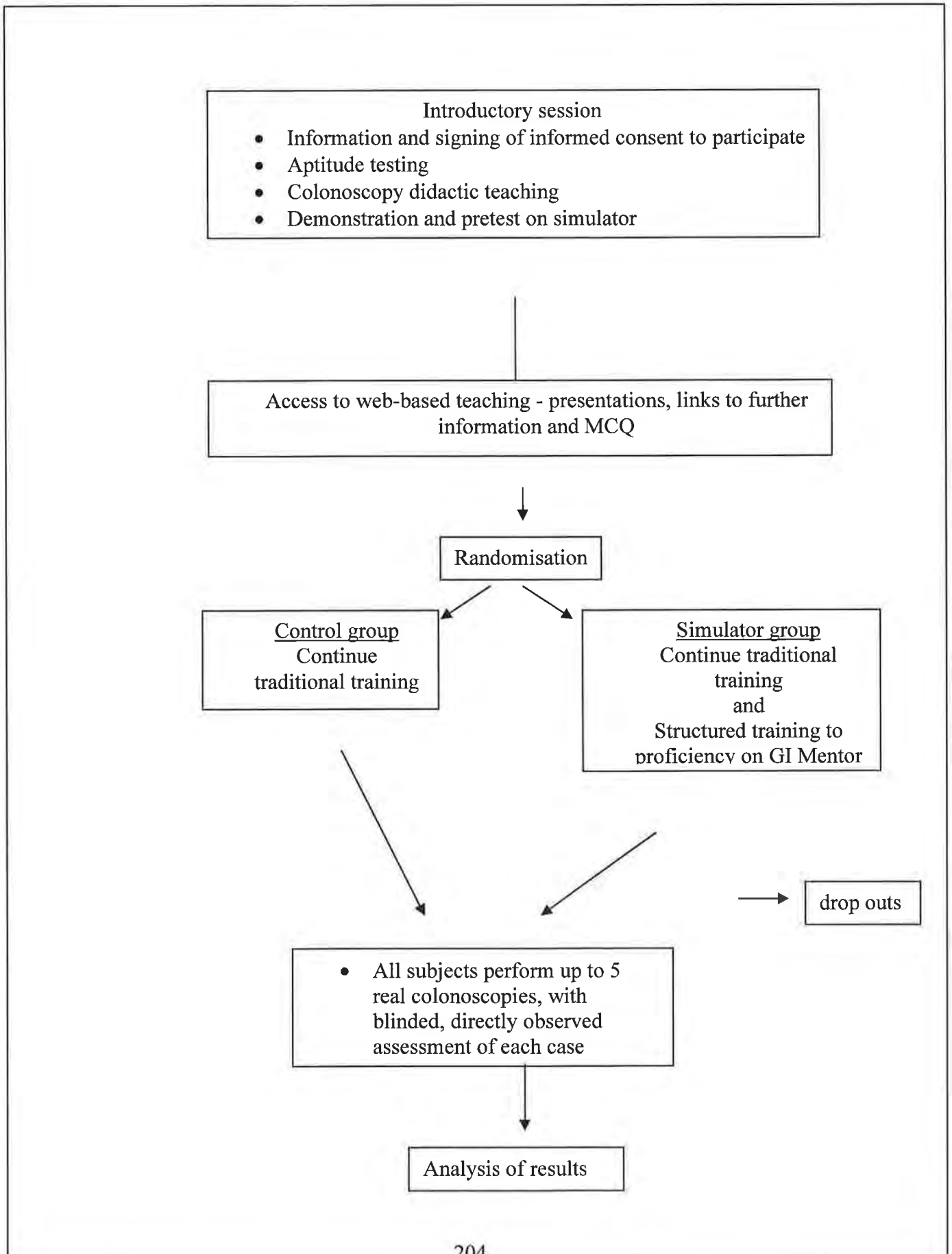
- Previous history of difficult colonoscopy (difficult anatomy, patient intolerance)
- Previous history of large bowel resection
- History of active inflammatory bowel disease, AV malformation or bleeding disorder
- Are expected to have a diagnosis of carcinoma confirmed by radiological imaging of any type.
- Currently are under the effect of anti-platelets or anti-coagulation therapy.
- Patients with ASA score \geq III
- Pregnancy
- Patient who cannot sign their own consent form (ie under 16 years of age, incapacitated by physical or mental disease)

7.2.4.3 Consultant evaluators

Experienced medical and surgical consultants currently in clinical practice who perform colonoscopies. As these consultants serve as trainers in the study, the same inclusion criteria were used as recommended by the Joint Advisory Group (JAG) on GI Endoscopy in the UK. Only surgical or medical consultants who have performed a minimum of 150 colonoscopies in the previous year were included in the study

7.2.5 Study design and procedures

Figure 7.1 – Overall study design



7.2.5.1 Introductory session

All recruited subjects attended an introductory session at the NSTC at RCSI in small groups. At the beginning of this session, the trial was explained in full and they had the chance to read subject information form and ask any questions, then every subject signed a consent form to participate in the trial. They also supplied demographic information. Other activities at the introductory session included:

Innate abilities testing

All subjects underwent a standard battery of perceptual, psychomotor and visuospatial tests as described for the colonoscopy study in chapter 2. The tests consisted of three “paper and pencil” tests, a computer based test and a test on the ProMIS surgical simulator (Haptica, Dublin). This combination of tests has been used at RCSI for several years for selection of candidates onto higher surgical training schemes. This was to ensure that no significant baseline differences existed between the groups which may have influenced performance. In addition, performance on these visuospatial tests has previously been shown to correlate with rate of learning on an endoscopy simulator.

Teaching and pre-test

This was delivered on every occasion by the same facilitators and consisted of a short lecture outlining the practical aspects of colonoscopy (consent, bowel preparation, sedation, endoscopic anatomy, colonoscope techniques for negotiating passage through the large bowel) accompanied by power point slides and a selection of video clips. The video clips demonstrated large bowel endoscopic anatomy, and

colonoscopic views as the scope is passed through different areas of the bowel. The teaching included a commentary detailing techniques on how to negotiate difficult areas

There was also a practical demonstration of a colonoscope using the simulator, with explanation of all the colonoscope controls, including techniques such as torque steering and resolving loops. Subjects also watched a simulated procedure (ie performed on the simulator) with commentary.

At the end of the session each subject performed a pre-test on the simulator, facilitated and proctored by one of the authors. The purpose of this was two-fold. Firstly we needed to have a baseline performance level which could be used for future comparison because despite satisfying inclusion and exclusion criteria there were still some baseline variations in ability. Some trainees for example had never even held an endoscope, whereas other trainees had performed some oesophagi-gastro duodenoscopies (OGD's) and parts of several colonoscopies. Secondly as there were subjects in both groups with no previous colonoscopy experience, those in the control group who had had no previous experience would have been unlikely to have been permitted to attempt any real colonoscopies unless they were at least familiar with how to hold the colonoscope.

7.2.5.2 Web-based learning

All trainees had access to teaching materials which were posted on a dedicated surgical education website. All trainees enrolled in surgical training programmes had access to the "school for surgeons" website. This is a Moodle-based web platform for delivering teaching materials. Trainees can access teaching information including presentations, pictures, videos, links to external websites etc. It also serves as a

discussion forum. Trainees who were recruited to the study had access to the colonoscopy training section. This contained sections on history of colonoscopy, technological advances, gastrointestinal anatomy, patient preparation, consent and sedation. There were links provided to websites of the British Society of Gastroenterology and the American Society of Gastroenterology in order to access published guidelines. At the end of the section there was a colonoscopy MCQ. Trainees were expected to achieve 100% on this test before being permitted to perform any real colonoscopies.

7.2.5.3 Randomisation

A simple block randomisation system was used to randomise subjects into the study or control groups. A block size of two was used as numbers were small and we wanted to keep the groups sizes as similar as possible. . In addition subjects were stratified for site and consultant to minimise the effect of possible confounding factors caused by site or consultant preferences. For example, a pair of trainees from one site would be randomised to control and study groups. In one case, OGDs performed was also used as a stratifying variable as we had 2 subjects from one site who had both previously performed OGDs

Subjects were randomised after enrolment and signing of consent which took place at the didactic session, in order to ensure similar numbers in both groups

1) Group CC (current curriculum); after the teaching programme as described above, subjects received no further structured training. They continued to have exposure to their conventional training in their respective hospitals. However in no cases did this result in a trainee completing a colonoscopy independently before the assessed cases.

2) Group SC (simulator curriculum); after the teaching programme as described above, subjects had supervised practice sessions on the colonoscopy simulator, as part of a proficiency-based progression curriculum

7.2.5.4 Proficiency-based simulator training

Training sessions were supervised on either a one to one or two to one basis. Practice sessions had a maximum time limit of 1 ½ hours although the average time was approx 1 hour. In addition to physically practicing the technical aspects of the procedure, subjects had to discuss the procedure preparation, and were given instruction on various aspects of colonoscopy during each session. In addition, after every procedure performance metrics were reviewed by the subject and facilitator. Subjects were also told about the target levels they were expected to reach, ie the proficiency level so that they had an idea of their progress and how far they were away from the target standard.

For initial practice sessions, case 1 (a simple diagnostic colonoscopy) was used, subsequently case 3 (a diagnostic colonoscopy with a tortuous sigmoid colon and difficult splenic flexure) was used. After their 2nd or third session, only case 3 were used. Trainees were logged in every time they had a practice session so their performance and learning curve was documented.

Proficiency level

The proficiency level was set from the mean performance of consultant gastroenterologists and colorectal surgeons. In total, 9 consultant surgeons and 5 consultant gastroenterologists contributed to the proficiency level. The metrics for this are outlined in table 7.1. All consultants performed Case 1, and then Case 3 twice in

order to acclimatise themselves to the simulator. Their mean performance results from their second attempt on Case 3 were then used to set the proficiency level. To reach proficiency level, subjects had to reach the targets for all the metrics in one performance. In addition, they had to reach this level on two procedures during one practice session but these did not necessarily have to be during two consecutive trials

Table 7.1 – proficiency levels for colonoscopy training

Metric	Proficiency level
% mucosal surface examined	86
time to reach caecum/s	488
% efficiency of screening	82
% time clear view lumen	92
excess local pressure (occurences)	1
% time patient in pain	0
looping (#)	0
lost view lumen	2

7.2.5.5 Assessment – real colonoscopies

All subjects in both groups performed colonoscopy procedures under full supervision by their consultant at their training centres. Subjects in the SC group were not permitted to perform colonoscopies until they had attained proficiency on the simulator as detailed above.

Regular, elective colonoscopy lists were used. The intention was that subjects would perform cases that were appropriate for trainees, and therefore only patients who satisfied the inclusion and exclusion criteria outlined in section 7.2.4.2 were eligible. Suitable patients were approached by the subject/consultant or member of the study

team. The study was explained and patients were given an information sheet to read. It was explained that participation in the study would not alter their procedure, that in teaching hospitals colonoscopies are usually carried out by medical/surgical trainees under senior supervision, and that their consultant would be present throughout the procedure and would take over if necessary, as is normally the case. Very few patients refused to take part in the study. If patients were willing to participate, they were asked to sign a written consent form.

Colonoscopy procedures were assessed by two evaluators; one was an independent, trained assessor who was one of the researchers for the study and had previously performed > 750 colonoscopies, and the other was the consultant who was supervising the endoscopy. Both evaluators were blinded to the subject's training status although they were aware that the trainee was involved in a colonoscopy training study.

The evaluators used a standardised assessment form. (Appendix 7.4) This incorporated dichotomous yes/no responses, and rating scales. Although the assessment form was developed for the purposes of this study it will incorporate elements from other validated assessment methods such as the global assessment method used by Park et al, JAG guidelines, and other credentialing guidelines. Both evaluators graded the procedures independently, and their scores were averaged to give a final score. Any serious complications which are not seen during the actual procedure ie a perforation were also to be recorded on the assessment form, but no serious complication occurred during the period of the study.

The consultant supervising the procedure was be instructed to treat the case as though it was a normal procedure for one of their trainees. They allowed the subject as much independence as possible while ensuring patient safety. If the subject was not making

progress for a fixed time period, the consultant took over the colonoscope and performed that part of the procedure, then returned the colonoscope to the subject if appropriate. At these parts, the supervisor noted if this was due to subject incompetence, or patient factors which made it unrealistically difficult for the subject such as an extremely tortuous or redundant sigmoid colon, or unexpected serious pathology which a junior trainee would not normally be expected to deal with. The take over was not regarded as an error where the case was felt to be particularly difficult.. The examination was considered complete when the cecum was intubated. To ensure a sufficient abnormality detection rate and in the patient's interests the consultant withdrew the colonoscope in all cases. The subject was be allowed up to 30 minute to reach the cecum and if they could not reach the cecum within this time frame, the consultant took over and completed the procedure.

If the consultant feels that patient safety was compromised at any stage, they took over the colonoscope and completed the case, as they would normally do. This decision was completely at their discretion. If this was due to an unreasonably difficult case, the subject was allowed to repeat the procedure on another suitable patient, and this subsequent procedure will be assessed as part of the trial. If however the consultant felt the case was a reasonable standard for the trainee, the case was scored taking into account their inability to complete the procedure and any other errors they may have made.

In order to gain additional information about the procedure, patient comfort levels were also assessed. This is one of the elements of colonoscopy assessment which has been recommended in the UK. This was judged by the supervising consultant and evaluator and recorded on the assessment form as part of the trainee evaluation.

7.2.5.6 Statistical analysis

Data were analysed using SPSS Version 15.0. Baseline differences between the groups were compared using Mann-Whitney-U tests as the numbers were small.

Correlations between simulator performance and real assessment were analysed using Spearman's rho correlation coefficient as different measurement scales were compared. For comparison of the checklist and global rating scores a larger number of observations were included which were graded on a numerical scale and Shapiro-Wilks tests confirmed normality of the data. Therefore independent T-tests were used to compare the performance of the control group and the study group.

7.3 Results

7.3.1 Overall results

24 subjects in total were recruited to the study, but only 17 completed their clinical cases.

Of the 7 subjects who did not complete the study:

- 3 subjects rotated to different hospitals during the course of the study and could not complete their cases
- 3 subjects dropped out through lack of interest
- 1 subject could not complete the cases because patient consent was refused.

Of the 17 subjects who performed clinical cases, 10 were randomised to the control group and 7 to the study group

64 colonoscopy procedures in total were assessed, but data was missing for 2 cases, leaving 62 which were analysed

Subjects were permitted to perform up to 5 colonoscopies each depending on availability of lists/patients/consultants/scheduling etc.

Demographic data are displayed in table 7.2

Aptitude abilities were assessed for all the subjects.

There were no significant differences between the mean aptitude scores when the subjects were grouped by gender, speciality, or group to which they were randomised.

Table 7.2 Characteristics of the subjects

	Current curriculum, n = 13	Simulator curriculum, n =11
Mean age	29.8	27.1
speciality	3 medical, 9 surgical	4 medical, 8 surgical
Previous colonoscopies	0	0
Previous OGD experience	6 yes, 7 no	7 yes, 4 no
Mean yr since graduation	2.8	2.77

7.3.2 Pre test results

All subjects completed one pretest on the simulator. This was partly to give a baseline against which improvements could be compared and also to ensure that no large differences existed between the groups. As the analysis was performed on an intention-to-treat basis, data for all subjects who enrolled in the study were included in the pre-test although some subjects dropped out of the study before performing any

endoscopies. The other purpose was to attempt to standardise the subjects as some had never held a colonoscope prior to the study and some had experience with OGD.

Pre test was performed using case 1 on the simulator. Subjects were given standardised instruction during the procedure, which to some extent standardised some of the qualitative metrics such as percentage of time with clear view, as they were instructed to keep the lumen in view.

However, time to reach the caecum was assessed. There was no significant difference in this metric when subjects in the training group were compare to those in the control group, but a large and significant difference when medical trainees were compare to surgical trainees. (Table 7.3)

A significant correlation was noted between the time to reach the caecum on the pre test and the overall assessment score for each subject’s first real colonoscopy.

Spearman’s rho -0.518, p = 0.033

Table 7.3 – pre test results for mean time to reach caecum

	medical	surgical	p-value (Mann Whitney-U)
Mean time to teach caecum/s	638	386	0.004
	Group 1 (cc)	Group 2 (sc)	p-value (Mann Whitney-U)
Mean time to teach cecum/s	470	464	0.994

7.3.3 Training results:

The mean number of cases required for the trainees in the simulator group to reach proficiency was 21 trials (sd 7.724). These were spread over a range of 2-5 sessions for every subject.

7.3.4 Assessment results.

In total 64 real colonoscopies were assessed, although assessment data were lost for 2 of these cases, so the analysis is based on 62 cases.

No serious complications were recorded in either group.

During the course of the study, 3 patients declined to give their consent to participate.

These cases were not assessed.

Table 7.4 – number of colonoscopies performed by every subject

subjects	Colonoscopies performed
6	5
4	4
6	3
1	2
Total - 17	Total - 64

Only one subject from each group would have reached the cecum independently without assistance. Subjects were specifically scored during the procedure for how much progress they were able to make without assistance (part of the checklist) There was no significant difference between the simulator and conventional curriculum groups in the scores for progress made.

The assessment form generates a checklist score, global assessment score and an overall score. Although all the scores were higher for colonoscopies performed by subjects in the simulator group, only the difference in score for the checklist reached statistical significance (Table 7.3, Figures 7.1 a-c)

Table 7.5 – assessment results

	Current curriculum	Simulator curriculum	p-value*
Checklist score	9.97	12.42	0.046
Global score	15.6	17.04	NS
Total score	25.6	29.5	NS

*Independent T-test, normality of data verified using Shapiro-Wilks test of normality

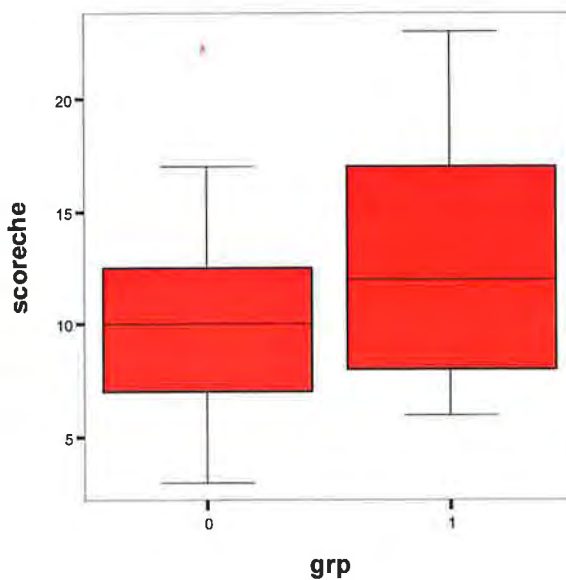


Figure 7.1a – boxplot showing the checklist scores for the 2 groups (0 = current curriculum, 1 = simulator curriculum)

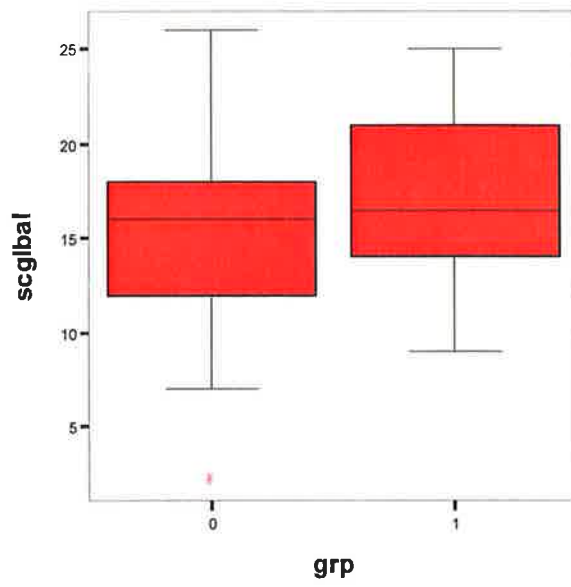


Figure 7.1b – boxplot showing the global scores for the 2 groups (0 = current curriculum, 1 = simulator curriculum)

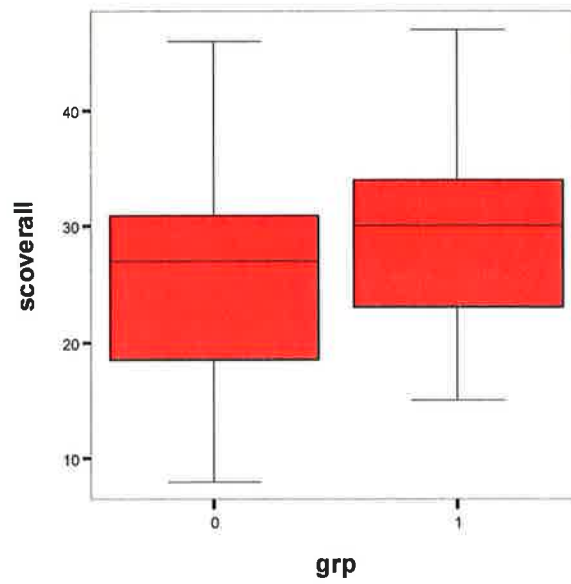


Figure 7.1c – boxplot showing the overall scores for the 2 groups (0 = current curriculum, 1 = simulator curriculum)

For all the individual metrics on the checklist part of the assessment form, the simulator group had higher scores, except for error score and ability to identify normal tissue. For these 2 metrics, both groups had the same mean score.

In the global assessment section, the simulator group also scored better for every metric except respect for tissue, where they scored lower than the control group.

The simulator group performed significantly better than the control groups for 2 metrics:

Ability to identify pathology: $p = 0.047$

Response to difficulties: $p = 0.043$

Although there was no significant difference between the two groups in the assessment score for progress made, we also assessed how far each subject would have progressed independently, ie if the consultant had not been available to take over. Table 7.4 shows the percentage of subjects in each group who progressed independently to particular locations, and figure 7.2 shows the same information.

Table 7.4 – percentage of subjects in each group who would have progressed independently to each landmark

location	Current curriculum	Simulator curriculum
Anus/rectum	89%	92%
Sigmoid/descending colon	36%	57%
Splenic flexure	25%	38%
Transverse colon	17%	19%
Descending colon	6%	8%
Cecum	3%	4%

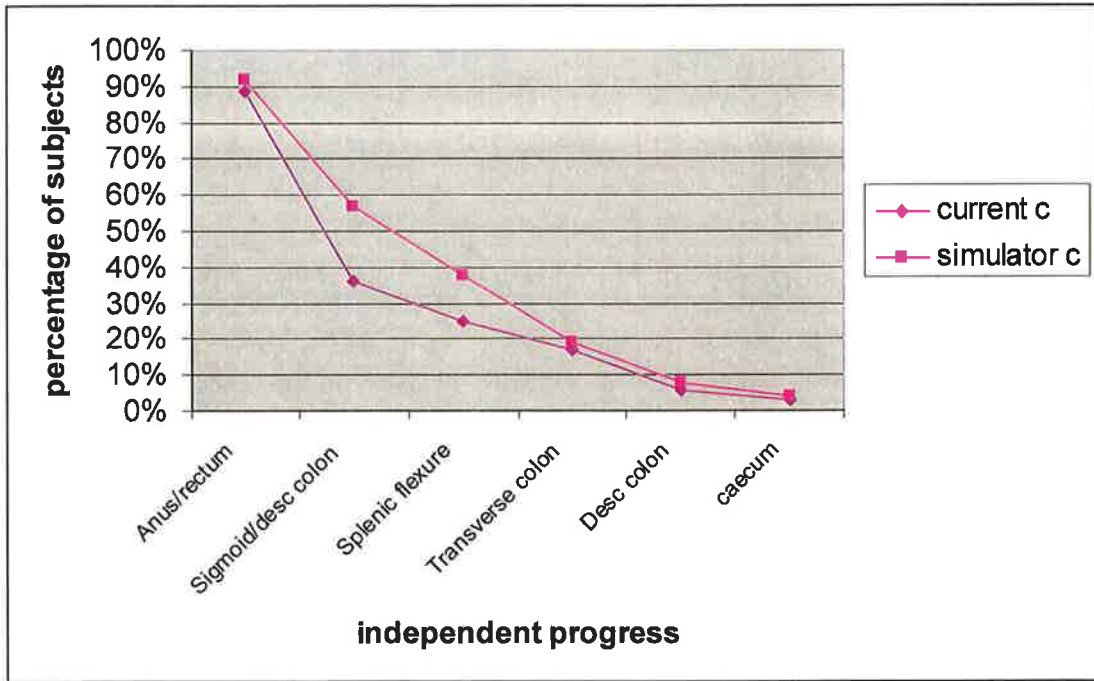


Figure 7.2 - – percentage of subjects in each group who would have progressed independently to each landmark

Although the subjects' performance improved over the course of the real colonoscopy cases (figure 7.3), there was no significant difference in the amount of improvement shown by the simulator group and control group.

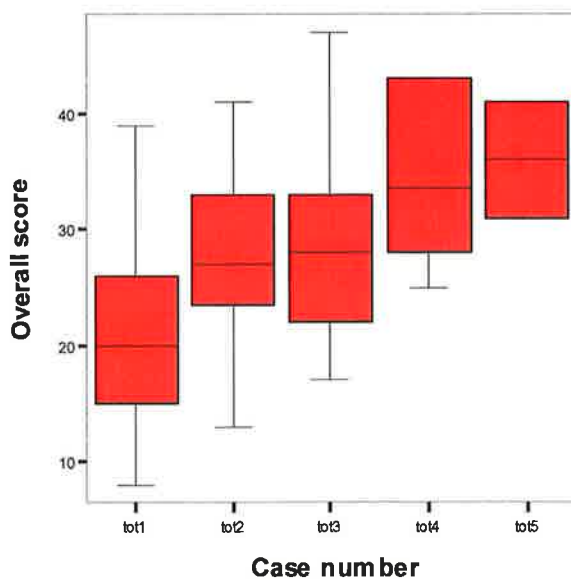


Figure 7.3 – boxplot showing the mean overall assessment score for all subjects for the real colonoscopy cases

7.3.5 Inter rater reliability

We assessed the inter rater reliability after the first 52 colonoscopies were assessed.

The inter rater reliability for the checklist part of the form was calculated using the formula: *number of agreements/number of possible agreements* and the mean score for the 52 assessments was 0.941, standard deviation 0.074.

For the global assessment part of the form, as it is marked on a continuous scale and the dataset was relatively small, a non-parametric correlation coefficient was used, and the correlation was significant at 0.84.

Given the high reliability of the assessments, subsequent colonoscopies were scored by one evaluator only.

Table 7.5 – interrater reliability using Spearman’s Rho

		Assessor 1	Assessor 2
Assessor 1	Correlation Coefficient	1.000	.838(**)
	Sig. (2-tailed)	.	.000
	N	52	52
Assessor 2	Correlation Coefficient	.838(**)	1.000
	Sig. (2-tailed)	.000	.
	N	52	52

** significant at $p < 0.01$ level

7.4 Discussion

The demand for colonoscopy has increased over the last number of years (BSG Endoscopy Committee, 2002). One reason for this is its introduction and use as a screening test for colorectal cancer in average risk subjects. In 2000, the American College of Gastroenterology recommended colonoscopy as the preferred colorectal cancer screening strategy, whenever the expertise, resources, and reimbursement for the procedure were available (Rex DK et, 2000). Colonoscopy has also been part of a national colorectal cancer screening programme in the UK since 2000 (Price J et al, 2005) and in Australia since 2006. There has been a recent pilot scheme in Ireland for screening colonoscopy and it is likely that this will soon be rolled out on a national level. (Board of the National Cancer Screening Service, 2008)

As the demand for colonoscopy has increased, so too has the need for high standards in colonoscopy performance. However it is known that quality of colonoscopy performance varies from centre to centre and endoscopist to endoscopist.

Colonoscopy is operator dependent and success of colonoscopy depends on the successful completion of the procedure. Given the hugely increased volume in demand for colonoscopy and the current (acgme.org) and proposed limitations in working hours for trainees (Dept of Health and Children, 2005), it is becoming increasingly difficult to train junior endoscopists in the traditional apprenticeship manner. The use of simulators to replace some of this training is a possible solution and previous research has demonstrated a benefit to simulator based training for endoscopy procedures (Tuggy ML 1998, Ferlitsch A, 2002, Park J et al, 2007) .

Simulator based training is safe, efficient and does not compromise patient comfort or safety.

Our secondary aim was to create and implement a colonoscopy teaching curriculum. A proposed template for developing a curriculum should include the following sequence: (1) didactic teaching of relevant knowledge (ie, anatomy, pathology, physiology); (2) instruction on the steps of the task or procedure; (3) defining and illustrating common errors; (4) test of all previous didactic information to insure the student understands all the cognitive skills before going to the technical skills training and in particular to be able to determine when they make an error; (5) technical skills training on the simulator; (6) provision of immediate (proximate) feedback when an error occurs; (7) provision summative (terminal) feedback at the completion of a trial; (8) iterate the skills training (repeated trials) while providing evidence at the end of each trial of progress (graphing the “learning curve”), with reference to a proficiency performance goal that the trainee is expected to attain. While the above is a proposed template, it includes in a stepwise fashion all components published in the literature that would comprise a comprehensive and validated training curriculum. All of the above details were incorporated in our curriculum design. Although feedback was not provided by an expert endoscopist, which may be optimal as discussed in chapter 4, feedback was immediate, consistent and focussed on performance goals and outcomes. In addition, feedback was provided in real time and not merely at the completion of each trial. Therefore this teaching curriculum can be implemented even in the absence of expert faculty.

Our results showed that the performance of colonoscopy by the subjects who were trained with a proficiency-based progression curriculum was significantly better when assessed with the checklist part of the assessment form. Although the difference in the

global assessment and overall ratings was not significant between the groups, the simulator group performed better than the conventional curriculum group.

This occurred despite the fact that our final numbers were smaller than had been planned. This was due to the relatively high number of subjects who did not complete the assessments, and the failure of all participating subjects to perform 5 colonoscopies each.

Although there was no significant difference seen between the groups for the learning curve over the first few colonoscopies, the performance of the simulator trained groups overall was better than that of the conventional group. Individually, the simulator-trained subjects performed significantly better for ability to identify abnormal tissue and response to difficulties, both of which are important skills in colonoscopy.

No differences in caecal intubation rates were reported between the groups. This is unsurprising as none of the subjects had performed a complete procedure before participating in the study and while caecal intubation rates are important markers of quality in colonoscopy (Rex DK et al, 2002, Bowles et al, 2007, the JAG) trainees do not reliably reach the caecum. This has been noted in previous studies such as the VR-OR study by Park et al 2007, which showed a significant advantage to simulator training but no difference in caecal intubation rates. However, when scored for how far the trainee would have been able to progress in the absence of assistance, a greater percentage of trainees in the simulator group were able to make more progress (Figure 7.3)

No difference was seen in serious complication rate between the groups. This is partly due to the low frequency of serious complications at colonoscopy (ie perforation rates of 0.07% to 0.12% have been reported, Iqbal CW et al 2008 and Luning TH et al

2007) but also due to the fact that trainees were closely supervised during the cases and would therefore have likely been prevented from committing any serious errors. Consultants were free to take over at any point they felt appropriate, due either to lack of progress or patient safety. This was an important ethical consideration in the study design and is a feature of most research involving assessment of actual patient cases. We encountered a number of logistical problems during the study. A lack of commitment by many of the subjects meant that some subjects who had been recruited did not complete any assessments, or failed to complete assessments before they changed rotation. As not all subjects at the time of participation were rotating through jobs with exposure to colonoscopy, scheduling difficulties and clashes with clinical commitments occasionally made it difficult to complete the clinical cases. As trainee cases took far longer than cases performed normally on the lists, we were limited with regard to how many cases could be accommodated on regular elective endoscopy lists. Another limitation was that subjects could only be assessed where there was direct supervision by the consultant, so not all colonoscopy lists were suitable as some were performed by other members of the team. Interestingly, patient refusal was encountered extremely rarely as an obstacle to the study and the vast majority of patients were happy to participate. Other difficulties which were hard to avoid were the baseline heterogeneity of the subjects. Although no subject had performed a complete colonoscopy prior to participation in the study, some subjects had never even held a colonoscope, while others had performed upper endoscopic procedures (a maximum of 10) and had performed parts of a colonoscopy before. However this is the realistic situation encountered in current training programmes with large varieties in training opportunities. We controlled for these differences by stratifying for previous

experience when subjects were randomised into different groups, but these baseline differences may have affected the results. In addition, there were no significant differences between the 2 groups for mean time to reach the cecum on the simulator pre-test, suggesting that subjects with varying baseline levels were probably evenly distributed throughout the 2 groups. Interestingly, surgical trainees appeared to have superior baseline skill level to medical trainees probably reflecting greater exposure to colonoscopy earlier in surgical training. While it would be preferable to recruit subjects with absolutely no endoscopic experience this would probably necessitate recruiting medical students as subjects and it would be unethical and unfeasible to allow medical students to perform complete colonoscopies. In addition, most consultants would be reluctant to allow trainees the chance to perform a complete colonoscopy if they had absolutely no previous experience and the aim of this study was not to compare simulator based training to no training, which would be a futile aim but rather to compare it to current training which is based on the apprenticeship method of trainees gradually being allowed to perform more of a particular procedure. Thus involvement of trainees in any kind of training-based research means assessing trainees who have varying baseline experience.

Another issue was the difficulty in controlling the amount of supervision provided by the supervising consultant. In order to endure patient safety, supervising consultants evaluators had to be permitted to take over the case when they thought appropriate but an inevitable consequence of this is the confounding effect caused by varieties in supervision style. We noted that different consultants had varying thresholds for taking over the case. Some would take over within a few seconds of the trainee encountering a difficult area, others would give the trainee several minutes to try to negotiate the area independently as long as there was no compromise to patient safety

and this undoubtedly biased some of the results. However, all consultants were given the same instructions prior to the commencement of any cases. Having 2 evaluators (for the first 52 cases) helped somewhat to minimise this bias, but if the trainee was permitted to carry out a small amount of the case only, they could only be assessed on what they had actually done.

Our inter rater reliability was extremely high for both portions of the assessment form. This was particularly true for the checklist section where we measured actual reliability or agreement and not just a correlation of the scores awarded by both evaluators. This makes the form a useful assessment tool which may be reliable even for one assessor.

Inclusion of both medical and surgical trainees was an important factor in the study. Endoscopic skills are important in both medicine and surgery and current training methods suggest greater supervision in medical training programmes than surgical. If surgeons are to maintain endoscopic standards at least equivalent to gastroenterologists it is important that training is provided at a sufficiently high level. As surgeons will always be called on to manage endoscopic complications, it is important that this area of expertise is maintained particularly in the era of laparoscopic surgery where precise tumour localisation is key. It is also relevant if surgeons are to retain a central role in the provision of Natural Orifice Transluminal Surgery (NOTES) (Wagner et al, 2008). The intensive, focussed training which can be delivered with the aid of simulation will likely be an important factor in the continued involvement of the surgical establishment in NOTES surgery.

7.5 Conclusion

We have demonstrated that a proficiency based training curriculum incorporating simulation can be effectively and efficiently delivered for colonoscopy. This provides a safe, standardised and efficient way of preparing trainees to perform colonoscopies in the clinical environment and trainees who have attained proficiency performed significantly better than controls on certain elements in the real clinical environment.

Chapter 8

Innate abilities and their relationship to surgical skills

Chapter 8 – Innate abilities and their relationship to surgical skills

8.1 Background

Given the increasing complexity of surgery and in particular the increasing role of image guided procedures and laparoscopic techniques, there has recently been interest in the existence of specific, innate abilities that may determine whether an individual will be able to master the different skill set required to operate in the 21st century. It has long been recognised that some individuals master the skill set required to operate laparoscopically in a relatively short space of time whereas others struggle to learn, take longer and may never fully reach competency.

Some research has focussed on the relationship of abilities in other manual skills to laparoscopic surgery, such as video game playing which was discussed in the previous chapter. Other activities which seem to be associated with better laparoscopic skills include musical abilities, sporting abilities, car driving (Nmura et al, 2008), and even prior experience with billiard playing (Stefanidis et al, 2006) and the use of chopsticks (Madan et al, 2008). However these are all associations only and do not imply causality.

A more reasonable explanation is the existence of innate aptitudes, encompassing areas such as visuospatial ability, perception and psychomotor coordination. Not all individuals possess the same amount of these innate fundamental abilities and those less endowed are likely to struggle during surgical training and thereafter in surgical practice. By definition these abilities are innate, or present from birth and cannot be altered by experience. They are relevant to the efficient performance of open surgical procedures but are probably even more relevant to minimally invasive surgery.

Generally accepted attributes fundamental to surgery include innate dexterity and technical skills (Cuschieri, 2001; Darzi, 2001) Neuropsychological factors such as

visuospatial organisation, stress tolerance and psychomotor abilities have also been shown to correlate with surgical skills (Scheuneman, 1984) Laparoscopic surgery requires surgeons to infer the shape of 3-D structures, such as the internal organs of patients, from 2-D displays on a video monitor, requiring adequate depth perception. Other challenges include adapting to the fulcrum effect, performing fine motor tasks at a distance from the operating field, inferring depth cues from shadows and light and adapting to a magnified image. It is clear that perceptual, visuospatial and psychomotor abilities are particularly relevant to laparoscopic surgery. The importance of such abilities is long recognised in other fields with similar requirements for manual dexterity such as the aviation industry. The existence of these abilities explains the relationship observed between abilities in related but different field such as music and surgery.

The significance of these abilities is clear. They could be used to select individuals most likely to excel in surgery and are arguably more relevant and objective than current selection techniques which rely largely on academic achievement (Gallagher et al, 2009; Cope et al, 2008; Bulstrode et al, 2003). Very few surgical training programmes measure these abilities when selecting candidates (Tansley et al, 2007). Results from tests could be used to counsel students about a career in surgery prior to the selection stage, or to counsel current surgical trainees regarding choice of specialty. This could also reduce drop-out rates from surgical training programmes which is wasteful and costly of time and money (Gallagher et al 2009; Schiven et al, 2004). Results from such tests could also be used to identify trainees who may take longer to learn specific techniques (Mc Clusky et al, 2005; Stefanidis et al, 2006) and who may require a longer training period or more intensive training.

There is debate therefore regarding the exact relationship of these abilities to surgery and their importance. Arguably if their only benefit is to predict rate of skill acquisition, their use as a selection tool may be unfair. The central tenet of proficiency-based progression curricula is that trainees can achieve objectively demonstrable standards of competency, and how many trials it takes to achieve this could be deemed irrelevant, particularly in the era of skills lab training, where significant portions of skills training can be removed from the clinical environment. In fact it has been suggested that while innate abilities themselves are unchanging, skills can be acquired through practice and they should arguably not be used to select surgical trainees but instead to guide different aspects of their training. (McCluskey DA et al, 2008)

Some groups have examined the relationship between novices' laparoscopic performance and innate abilities. Schiven et al demonstrated concurrent validity in aptitude testing, by comparing scores of surgical novices on a laparoscopic cholecystectomy virtual-reality simulation with performance scores on a battery of standardized psychometric aptitude tests. The abstract reasoning and the space-relation test had predictive and selective value, identifying individuals who have good laparoscopic surgical virtual-reality performance. Hassan et al showed that among novices, visual-spatial perception (as measured by standardised tests) is associated with manual skills performed on a virtual reality simulator. Gallagher et al (2003) showed a correlation between psychometric perceptual scores and novices' performance on a laparoscopic cutting task. However, these studies have examined novices' baseline abilities only, with no indicator of how quickly novices will improve their performance.

Rates of skill acquisition have been examined to address this uncertainty. Mc Clusky et al (2005) found a correlation between some but not all innate abilities. They demonstrated that visuo-spatial ability did not significantly correlate with training duration for a simulated laparoscopic task, but was significantly related to perceptual and psychomotor aptitude. Ritter et al found a stronger association, and demonstrated significant correlations between visuospatial, perceptual and psychomotor abilities and time required to reach proficiency on an endoscopy simulator. Stefanidis et al (2006) found that psychomotor testing had limited use for predicting baseline performance, but was a predictor of the rapidity of skill acquisition in a laparoscopic simulation curriculum.

All of these studies have examined the relationship of innate abilities to performance in laparoscopic or endoscopic tasks. There is less evidence demonstrating a relationship with open surgical skills. Dashfield et al (2001) performed a small study looking at knot-tying, and found that psychomotor abilities correlated with trainees' initial proficiency in learning to tie a surgical reef knot, ie the improvement between their first and second knot, with training provided between the two attempts.

Clearly there is much evidence supporting a link between innate abilities and laparoscopic /endoscopic skills, but the exact nature and significance of the link requires further definition. Most studies mentioned here are small, and almost all have examined novice performance or improvement only.

To further address these questions, we measured innate abilities of all the subjects who participated in the four previously described studies, with a partial data set only for the HALC subjects. If these abilities are significantly linked to performance, optimisation of surgical training should include the concept of selection into surgical training in order to train those who are most likely to benefit.

8.2 Methods

Aptitude tests were used as described in chapter 2. Three main aptitudes were measured – visuospatial, perceptual and psychomotor ability.

To investigate the relationship, we measured each of the 3 aptitude tests as follows:

Visuospatial tests – each of the 3 paper tests is marked out of a different total. To obtain an overall figure, each score was converted to a percentage. Percentage score for each individual test were correlated with the performance measures.

Perceptual – this test generates a single score between 0 and 1

Psychomotor – this test generates 3 results: time (T), instrument path length (IPL) and instrument smoothness (IS)

For all correlations, non-parametric correlation coefficients (Spearman's Rho) are used unless stated otherwise.

All performance data and testing were measured as described in chapter 2 and 3-6

8.3 Results

Fundamental abilities for each of the 4 main data sets were analysed separately.

8.31 Endovascular subjects

Each of the 3 aptitude results were correlated with 7 of the principle outcome measures on the VIST simulator (time, contrast use, fluoroscopic time, errors, balloon placement accuracy, residual stenosis and percentage of lesion covered.)

As all subjects performed 6 trials each the following comparisons were made:

- Correlation between aptitude and performance on trial 1

- Correlation between aptitude and performance on trial 6
- Correlation between aptitude and summed performance measures over the 6 trials
- Correlation between aptitude and percentage difference in performance between trial 1 and trial 6

For each of these calculations, p values of < 0.05 were taken as significant.

Significant correlations only are presented here.

Given the above calculations, (5 aptitude scores x 7 outcome measures x 4 comparisons) 140 results were generated in total and only 7 of these reached statistical significance. These are outlined in table 1 and as can be seen none of them reached a value of 0.8.

We also correlated the results with the individual tests which make up the visuospatial tests and again found no significant results

Interestingly, 6 of the 7 significant scores were for the IPL measurement on the psychomotor task suggesting that this may be the most valuable metric to score.

Within the aptitude tests themselves, PicSO_r was significantly correlated with the visuospatial tests.

Although relatively few results reached statistical significance, the number of subjects analysed was small at 18, and a bigger group of subjects may well have yielded more significant results.

Table 8.1– summary of relationship between fundamental abilities and performance presented as correlations. Statistically significant correlations only are presented

period	metric	visuo spatial	perceptual	psychomotor			p-value
				T	IPL	IS	
Trial 1	fluoro time				-0.542		0.025
Trial 6	placement accuracy/mm				0.574		0.016
	residual stenosis				0.755		0.000
	lesion coverage/%				-0.759		0.000
Total 6 trials	Placement accuracy				0.624		0.040
	Residual stenosis		0.755				0.000
Improvement 1 to 6	placement accuracy/mm				0.549		0.022

8.3.2 Colonoscopy subjects

All the results of the aptitude tests were correlated with the pretest performance data to assess for any differences at baseline.

The 3 tests (ie visuospatial, perceptual and psychomotor) were correlated with the results for pretest time, pretest global score, and number of attempts required to reach proficiency in the training group. Out of these correlations the only results to reach statistical significance were for the correlations between:

Pretest time and psychomotor time

Overall pretest score and psychomotor instrument path length.

Therefore it could be presumed that the psychomotor attributes of subjects are the most closely related to ability in endoscopic performance.

Table 8.2 – relationship between aptitude tests and pretest performance presented as correlation coefficients. Statistically significant results only are presented

metric	visuo spatial	perceptual	psychomotor			p-value
			T	IPL	IS	
Pretest time			0.472			0.036
Overall pretest score				-0.544		0.024

We then assessed the relationship between aptitude and performance on the real colonoscopy task. We correlated all the same innate ability results with the performance assessments on the first real colonoscopy.

All the correlations between the visuospatial and perceptual results and colonoscopy performance results were positive, and those between the psychomotor results and colonoscopy performance results were negative, ie a longer time on the psychomotor task was negatively correlated with performance. However, the only correlation to reach statistical significance was between the score on the perceptual test and the global performance score on the first colonoscopy.

Table 8.3 – correlation between actual colonoscopy performance and innate aptitudes

metric	visuo spatial	perceptual	psychomotor			p-value
			T	IPL	IS	
Global score first colonoscopy		0.509				0.044

8.3.3 Wii subjects

We correlated the aptitude results with performance in session 1 and session 2 and a similar pattern was observed. Of all the correlations performed, few reached statistical significance and these results only are displayed in tables 8.4 and 8.5

Table 8.4 – correlations between aptitude tests results and bead transfer task

Session	metric	visuo spatial	perceptual	psychomotor		
				T	IPL	IS
Session 1	Bead - IS			0.499 (0.03)	0.506 (0.027)	
	Bead - time			0.559 (0.013)	0.522 (0.02)	
Session 2	Bead - time				0.505 (0.033)	

Table 8.5 – correlations between aptitude test results and cutting task

Session	metric	visuo spatial	perceptual	psychomotor		
				T	IPL	IS
Session 1	Cutting - time			0.505 (0.027)		
	Cutting - IPL			0.469 (0.043)		

None of the correlations were strong – as can be seen from table 8.4, the highest correlation was 0.559. All of the significant correlations occurred with the psychomotor tests only and none of the results for visuospatial or perceptual tests correlated significantly with the data.

The % improvement from session 1 to session 2 did not significantly correlate with any of the aptitude results.

Interestingly, there were 7 significant correlations in total, and 6 of these occurred with performance results from session 1. This suggests that psychomotor ability may predict a better laparoscopic performance in novices at baseline, but that after even one practice session, these abilities no longer predict improved performance, indicating that any baseline differences in ability are quickly overcome by practising. As discussed in chapter 5, the improvements in this group from session 1 to session 2 were large.

8.3.4 HALC subjects

For this group, less data were recorded. Innate abilities only were measured for group 2, the feedback group, and within this group, visuospatial tests only were conducted. Within this group, data was missing on two of the subjects, meaning that visuospatial ability only was measured for 10 of the subjects.

Despite these small numbers, some significant results were obtained. Within the visuospatial tests, the results for the Card Rotation test (CRT) were significantly correlated with overall IPL and IS for all 5 trials. There was no significant correlation with error scores, and neither of the other two visuospatial tests (Cube Comparison and Map Planning Tests) correlated with any of the performance measures.

Table 8.6 Correlations between innate abilities and HALC performance

Visuospatial test	Performance measure	correlation	p-value
CRT	IPL (overall)	-0.806	0.005
CRT	IS (overall)	-0.733	0.016
CRT	IPL (trial 1)	-0.806	0.005
CRT	IS (trial 1)	-0.697	0.025
CRT	IPL (trial 5)	-0.66	0.038
CRT	IS (trial 5)	-0.527	NS

As can be seen from Table 8.6, the correlations were strong and significant for the overall results and results for trial 1. For trial 5, the correlation was significant with IPL only, and this was less strong, suggesting that the performance benefit resulting from superior visuospatial ability may lessen with practice.

8.4 Discussion

To summarise the findings, our results showed several significant correlations between innate ability and performance data across a range of 4 different types of surgical procedures, ie hand-assisted colectomy, endovascular, endoscopic and basic

laparoscopic procedures, although the correlations were weak. It is interesting to note that while none of our data sets showed any significant correlations between surgical performance and all the innate abilities taken together, individual correlations were seen between different aspects of the innate ability tests and performance.

Visuospatial ability did not correlate with any performance outcomes for 3 of the experiments, yet one of the tests correlated strongly with performance outcomes for the HALC data. Perceptual ability correlated (weakly) with only one performance measure across all 4 experiments. The remainder of the significant correlations were with psychomotor ability only, suggesting these abilities are more relevant to ability in image guided procedures. It is clear from our data that the innate abilities taken together do not appear to predict improved performance.

It seems intuitive that the psychomotor abilities seemed to be more relevant than the other tests. This ability in particular seems most relevant to surgery, and related to the well known expression “good with his hands” However, our results may be a factor of the way we measured this ability. It was measured on a laparoscopic surgical simulator. Arguably therefore, subjects with more surgical experience may do better on this test than those with less and therefore this psychomotor test may be affected by prior surgical ability as a confounding factor and may in fact be measuring this rather than innate psychomotor ability. We found this previously in other work done in our unit, where psychomotor ability as measured on the ProMIS surgical simulator varied significantly between groups with different surgical ability.

Interestingly, the card rotation test (CRT) correlated with several performance outcomes for the HALC procedure. A correlation between visuospatial ability in particular and simulator performance has been previously demonstrated by Hassan et

al 2005, although a different battery of tests was used. Our results suggest that ability to mentally manipulate 2-d figures is relevant to laparoscopic skill.

The exact role of innate abilities testing in surgery is still unclear. However, it is clearly relevant. Testing is mandatory in other high stakes industries such as aviation and the military. It may be that current methods require additional refinement. The Royal Air Force (RAF) for example uses innate abilities in a domain centred way. In other words, relevant domains are identified for each of the different roles they recruit for such as air crew, personnel manager etc, and only abilities pertaining to these domains are measured as part of suitability testing (Tsang et al, 2003) Even in the literature pertaining to innate abilities and surgery a similar pattern can be observed, ie only certain innate abilities appear to correlate with surgical performance. (Shaven et al, 2004, McCluskey et al 2005, Gallagher et al 2003)

There are many other abilities which can be measured by psychometric testing which are also relevant to a surgical career such as personality traits. While manual dexterity in surgery is an important trait, it is not the only quality necessary for a successful career. Selecting trainees on the basis of these abilities therefore may be fair and objective but may not be appropriate. There does not appear to be a relationship between these abilities and the career aspirations of junior doctors. (Cope et al, 2008)

The exact role of these abilities may lie more in the arena of guiding training and identifying particular trainees who may need more practice in a particular area such as fine laparoscopic dissection or left handed dexterity. (Mc Cluskey et al, 2008)

Therefore measuring these abilities is an important way in which training can be optimised as curricula could then be created around the specific needs of different trainees. This is a process that is clearly only feasible within the context of a skills lab, but in this environment it would be straightforward to design. Trainees could undergo

batteries of tests at the commencement of training and then follow individualised training plans based on this information. This is certainly a potential future direction for surgical training and should make training more individualised and efficient.

8.5 Conclusions

We found significant correlations between individual innate abilities and surgical performance although not for all innate abilities taken together. While the exact role of these abilities in surgery requires more definition, they could be used as valuable tool to guide and optimise surgical training

Chapter 9

Conclusions

Chapter 9 Conclusions

9.1 Overview

Surgical training is at a turning point. Working hours for junior hospital doctors have been one of the most-discussed issues in medicine in Ireland over the past 6 months and have been the subject of High Court actions (<http://www.bailii.org/ie/cases/IEHC/>). The European Working Time Directive limits on working hours, which specify that no more than 48 hours can be worked in a week, are due to be put into effect from October 2009 and shift work is set to become a reality. Similar changes have taken place in the UK and the USA (Carlin AM et al 2007, Feanny MA et al, 2008). At the same time, surgery is becoming more and more sophisticated. Increasing specialisation and technological advances are just two of the factors which are impacting on training. For example, robotic assisted radical prostatectomy was introduced to Ireland in 2007 and 30 cases have been performed to date (P O'Malley et al 2009). The field of minimally invasive surgery continues to develop, and future challenges include Natural Orifice Transluminal Surgery (NOTES) and Single Incision Laparoscopic Surgery (SILS) and intensive training will be required before these modalities can enter clinical practice.

Even in the last few years, surgical training has developed to adapt to these challenges. Skills lab and surgical simulators are becoming more and more common and the benefits of simulator training in particular are now accepted. Mounting evidence supporting the relevance of this training to real operative performance (Seymour N, 2008, Gurusamy K et al, 2008) has increased this acceptance. The last Clinical Congress held by the American College of Surgeons in October 2008 for

example included a forum on how to set up and run a skills lab. Although expensive, with increasing demands and more focus on competency assessment and patient safety, surgical simulators are increasingly available and accessible and look set to become as integral to surgical training as flight simulators are to pilot training. However, the availability of facilities does not ensure their optimum use and there is still a growing need to maximise the benefit obtained. The aim of the work presented here was to investigate ways to optimise surgical training, principally using simulators as a medium, but also to determine other factors that may be relevant to efficient and effective training.

9.2 Principal findings of the studies

The data for this research was generated from 4 studies. All studies incorporated simulator-based training and one also had a clinical component. The main procedures trained in the studies were Hand-Assisted Laparoscopic Colectomy (HALC), endovascular Renal Angioplasty and Stenting (RAS), colonoscopy, and laparoscopic basic tasks. We also used a virtual laparoscopic cholecystectomy procedure.

All subjects recruited to the studies were novices in the index procedure for the study, although with the exception of the subjects in chapter 5, all were surgical trainees with varying levels of expertise.

All studies showed a significant improvement in performance, even across a relatively brief practice schedule. While this was not the specific aim of any of the studies, it clearly highlights the advantages of controlled, non-clinical based training, where specific procedures or procedural steps can be repeated continuously until performance is improved (Kaufman 1987, Tesch-Romer 1993).

All of the studies except chapter 5 incorporated cognitive training. This is an important part of any skills training programme (Van Herzele et al, 2008, Gallagher et al 2005.) Trainees practice to become competent surgeons, not skilled technicians. In addition to making the training more clinically relevant, the cognitive training increases the benefit accrued from the simulator based training and explains why almost all trainees were able to complete full procedures within their first 1-2 attempts.

9.2.1 Feedback

One major element we investigated was the benefit of providing performance feedback. If there is no benefit to this, trainees could practice independently in the absence of facilitators or teachers which would make training logistically easier to organise. Chapters 3 and 4 investigated this and similar findings were seen – while trainees improved their generic procedural skills such as instrument handling in the absence of feedback, procedure specific errors scores were reduced with standardised feedback. The relevance of this is slightly different in the two chapters. For HALC procedures, the improvement in metrics such as instrument path length has less practical significance in the real environment although it may reduce procedural time, and can be seen as a marker of superior skill. However, improvements in metrics such as fluoroscopic time are relevant to endovascular procedure where minimising the patients' exposure to radiation is clinically relevant. The reduction in errors is significant to both procedures, as many of the errors assessed would be causes of significant morbidity in real life, such as colon perforation, poor anastomosis and ureteric injury in the HALC study, and renal perforation or arterial injury for the RAS procedure. In addition, chapter 4 investigated the influence of feedback provided by

experts and non-experts. Although objectively assessed error scores were lower in the expert group, the difference between group B and C did not reach statistical significance, suggesting that non-expert feedback may be equally beneficial to expert feedback. However, an additional factor is that trainees may benefit more in the presence of an expert due to increasing pressure to perform well which may confer an additional benefit although this was not clear from our data.

9.2.2 Video game training

Despite the existence of much published evidence suggesting superior surgical ability in those who play video games, our study did not show a significant performance improvement following a short period of video-game training in a prospectively assigned group of subjects except for one performance metric. As discussed this was most likely due to an insufficient practice schedule, although it is useful to be able to define minimal periods of practice which may be of benefit. We can then infer that 3 hours practice is insufficient, and perhaps a 2-weekly period would be of more benefit. However, the trend was towards an advantage to video-game playing, and the additional advantages are that this type of training is cheap, easy, does not require trainer presence or specialised facilities and can be regarded as recreational. One area in which this research may be particularly useful is in developing countries where financial and other restraints would not support advanced skills labs or simulators.

9.2.3 Skill retention

Despite massed practice sessions and no intervening reinforcement, we found evidence of skill retention. Subjects performing the HALC procedure demonstrated complete retention for error scores, highlighting the benefit of proficiency based

training as this was the performance element for which proficiency was obtained. Benchmark levels should thus be incorporated into every skills training programme in order to maximise skills retention. Subjects who performed RAS procedures also demonstrated considerable although not complete retention at a mean period of 4 months, likely due to the intensive training schedule.

9.2.4 Real-world benefit of a structured training curriculum for endoscopic novices.

Trainees who underwent a structured, proficiency based training curriculum for colonoscopy demonstrated an improved clinical performance when compared to control subjects when assessed with a structured checklist. Their global assessment score showed an advantage to the simulator curriculum although the difference did not reach statistical significance. The curriculum incorporated didactic and web-based teaching, supervised practice sessions with proximate feedback and a requirement to pass a short test of knowledge before commencing real colonoscopies. This provides a template for the design of similar skills courses and could be applied to other surgical procedures. Importantly, the subjects in the simulator curriculum only required 21 repetitions to reach proficiency which does not require a substantial time commitment. Another relevant point was the inclusion of both medical and surgical trainees, demonstrating the contribution that structured training can make towards procedure-driven rather than speciality-driven practice. This was despite significant inter-speciality differences at baseline testing. Curricula such as we designed may therefore be of particular benefit to future cross speciality procedures such as NOTES, which will require the expertise of both surgeons and gastroenterologists.

9.2.5 High inter-rater reliability of assessments

We specifically examined inter-rater reliability of our assessment methods for the HALC and colonoscopy procedures. We found extremely high inter rater reliability, demonstrating the robustness and validity of the assessment that we designed.

Assessment is a vital part of training and it should be standardised and objective. The forms that we designed could potentially be used for future skills courses and assessments. In particular we assessed interrater agreement for our checklist forms and not merely correlations, and demonstrated agreements of 0.941 for the colonoscopy data and 0.969 for the HALC assessments, both of which were highly significant. Our global rating correlation was also measured at 0.84, which is high for this type of assessment. We used two modalities in our assessments – a checklist section with dichotomous yes/no responses to measure specific elements and a global rating section which can allow for greater discrimination between subjects which may not be picked up with the checklist assessment. A combination of both elements has been used by many other research groups for the same reason. (Reznick et al, 1997)

Our checklist in particular had very tightly defined errors which increased the interrater reliability. In addition, raters underwent training before commencing any assessment, and we suggest this should be a mandatory requirement for any examiners or assessors.

9.2.6 Influence of innate abilities on surgical performance

We demonstrated some correlations between specific aspects of performance and individual components of innate abilities, but few overall results. In general, psychomotor ability appeared to correlate most frequently with performance although as discussed this may be a factor of our assessment method. Our individual data sets

were small which may have impaired our results, but the correlations we observed, while small in number were significant. We believe continued measurement of these abilities is important but that its value will lie in identifying areas of weakness in different trainees, facilitating the design of specific curricula around the needs of individuals, which would represent optimisation of surgical training at a personal level.

9.2.7 Skills transfer

We investigated transfer of skills between different procedures in the sub study described in chapter 3. It would be beneficial if training to proficiency on one procedure would improve skills on a different procedure, as the training would have more value. We assessed skills transfer between the simulated HALC procedure and virtual colonoscopy and laparoscopic cholecystectomy procedures. We found no significant transfer of skills to the colonoscopy procedure, but subjects did appear to improve their performance of the laparoscopic cholecystectomy after completing the HALC training. However this does not imply causality and may just reflect the independent learning curve for the virtual laparoscopic cholecystectomy itself. Factors which may have affected the results included trainee fatigue and the fact that they were not required to reach proficiency on the HALC procedure. Inter procedural skills transfer has not been widely demonstrated in the published literature, but is an area that may warrant further research.

9.2.8 Self-assessment

Self-assessment is a valuable part of training and should be incorporated into every training curriculum. We incorporated self-assessment as part of the feedback process so that trainees would reflect on their own performance before being given objective feedback. Overall, self-assessment mirrored performance closely although the relationship was less clear for individual metrics. This may be due to differences in accuracy of performance assessment among different trainees (Chapter 3). Some trainees reliably assessed their own performance and others were less accurate. If trainees could self-assess their performance with greater precision it may reduce the necessity for faculty to be present at training sessions.

9.2.9 Feasibility of simulator training

This is an important aspect of any curriculum. Work-hour limitations mean that training must be efficient and compatible with clinical duties. In all our studies, training took place over brief periods of time and did not interfere with any of the trainees' clinical duties. With the exception of the HALC training which involved consumables such as the instruments and anatomy trays, the simulator training is low-cost once the simulator has been purchased initially. The trainees all gave positive feedback regarding the training. In chapter 3 for example, trainees felt that they had benefitted from the training and agreed that they would re-attend a similar course. Trainee motivation is an important aspect for any curriculum and enhances the educational experience.

9.3 Limitations of the studies

In the large and expanding area of surgical training and education, this data is limited to examining certain aspects only, and inevitably there are limitations to the studies. The first is the small number of subjects per study. Although many of our results were significant, individual data sets were small. There were many reasons for this, including funding of research materials, capacity of our skills lab, (which is the National Surgical Training Centre and must cater for the country's trainees) and availability of trainees who have busy clinical jobs. This particularly affected the colonoscopy study, where the actual number of clinical cases assessed fell short of projected numbers due to trainee availability and drop-outs, and patient availability. Only one of our studies had a clinical correlation, but this was not the aim of the other studies which specifically addressed training through simulation. There is an abundance of published evidence supporting VR to OR, and adding data to this was not the purpose of this research.

By virtue of their design, our studies focussed on skills training. Technical ability is an important aspect of surgery, but not the only aspect, and there are many other areas relevant to surgical training which we did not examine. This fell outside the scope of this thesis, but we acknowledge that we have addressed a subset of the elements which are relevant to surgical training.

All our studies involved trainees who were novices in the particular procedure, although had varying skills levels. Skills training is not solely relevant to novices but also to those at intermediate and advanced levels, especially for credentialing and continuous competency assessment, and therefore we have only examined training in one of the experience cohorts for whom training is a relevant issue.

9.4 Future uses of the data

There are many applications of our research findings, mostly relating to the design of future skills courses and curricula. Intensive courses may be less effective than distributed training, but are still of value and novices can expect to see significant improvements even after a half-day training. This is relevant in an era of reduced working hours where less time is available for training. As most hospitals do not possess individual skills training labs, trainees generally have to travel to attend skills courses so it is reassuring to see that there is a benefit even when the time available to attend such courses is limited. Absolute novices can also benefit from such courses and this may particularly be the case when there is a cognitive teaching session in addition to the practical session. Availability of course materials online in advance may be another way of optimising the benefit of intensive courses. Ideally proximate feedback should be provided although this does not have to be provided by an expert, and for some outcomes may be equally useful when provided by a trained facilitator. It may thus be very acceptable to design courses even in the absence of clinical experts to provide teaching, provided that non-clinical facilitators are sufficiently familiar with the simulator and procedure. Trainees can be reassured that this method of delivering training will be beneficial. Another way of designing a course with these research findings in mind would be to structure trainee unsupervised practice initially, where trainees would have the opportunity to improve instrument handling skills, and then have further training sessions with a trainer which could then focus on procedural specific techniques and error avoidance. This may maximise the benefit of the clinical expert.

Trainee self assessment appears to reflect actual performance and may be a useful adjunct to training as even the process of self assessment forces trainees to examine

their performance. This process has already been commenced in some training schemes. For example, the basic surgical training (BST) scheme administered by RCSI requires trainees to intermittently self-assess their performance in their post as part of the CAPA process. This is a valuable exercise but the process of self-assessment could be extended and should be a more integral part of the training programme. It could be incorporated into logbook completion, whereby a trainee would have to provide a brief self- assessment of every procedure that they perform which is entered into their logbooks. The simulator can facilitate the process of self-assessment as it records video footage of every performance.

Even when procedures which are taught are not practised by the trainee for several months, we have demonstrated partial to complete retention of skills. The effects of skill decay may therefore be less than previously thought but our data suggest that attaining proficiency in a task reduces the amount of knowledge decay when the procedure is not rehearsed. This is a real issue given current training rotations where a trainee may not be exposed to a certain element of surgery for several months, and skills they had previously obtained may not persist. Access to a simulator even for brief periods of time may be sufficient to offset this process and help to refresh skills which had been previously acquired.

Training courses should include a cognitive training element which addresses clinical applications. This makes the procedure more relevant to trainees and will increase their interest and their willingness to behave the same way in the simulated environment that they would in the real environment, thus maximising the benefits obtained from the training. Training in other skills involving manual dexterity may well benefit surgical performance and does not detract from training hours.

Trainee innate ability may be usefully measured in order to direct training and we have identified aspects of fundamental ability which may be especially relevant to surgical performance. However, as discussed in Chapter 8 it seems difficult at this point and in the absence of larger data sets to justify performance in innate abilities tests as a selection tool for entry onto surgical training programmes. As discussed there may be more relevance in the identification of domains of surgical ability which some individuals possess in greater quantities. An obvious example here is the difference between breast surgery and colorectal surgery for example. The former relies almost exclusively on open surgical techniques while the latter is heavily reliant on laparoscopic surgery. The identification and accurate measurement of different domains of psychomotor ability may help to suggest which trainees are more suited to which branch of surgery or alternatively to counsel trainees who wish to pursue a particular branch of surgery that they are likely to require longer training periods in order to become proficient.

We have incorporated many of these elements to design a structured curriculum for colonoscopy, which was easy to administer, efficient and catered for medical and surgical trainees. Colonoscopy is likely to become a procedural driven rather than a speciality driven area. This makes it even more relevant to train endoscopists from different backgrounds to the same level of competence, and this curriculum caters for trainees from both medical and surgical backgrounds. The advent of screening colonoscopies which will shortly be introduced in Ireland will force endoscopy training to become more efficient, available and standardised. Another factor is the future development of NOTES procedures which will demand expertise in endoscopic techniques and likely require simulation training before clinical practice. The demand for simulator training for colonoscopy is thus likely to increase and should be part of a

structured curriculum. Training to proficiency with this curriculum improved real clinical performance and this could be used as a template for many other skills courses. Trainees would access didactic teaching materials,, pass a cognitive proficiency test, then have supervised practice sessions until they reach a predefined standard for any one of a number of index procedures. This would be likely to lead to improved clinical performance.

9.5 Final conclusion

We have identified many ways in which surgical training, particularly simulation-based training can be optimised. Much future training is likely to be based in the non-clinical environment of skills labs and simulators and in the current climate of changes and challenges to surgical training it is imperative that the benefits of this training are maximised. Surgical training must be efficient, standardised and high-yield and it is to be hoped that current pressures on training will lead to an improved system for current and future trainees. We believe that the findings of this research are thus topical and relevant and that they may be used to continue to improve surgical training. Better training will produce better surgeons and ultimately better patient care which is the primary goal.

References

References

1. Aalbers AG, Biere SS, van Berge Henegouwen MI, Bemelman WA. Hand-assisted or laparoscopic-assisted approach in colorectal surgery: a systematic review and meta-analysis *Surg Endosc* 2008; 22(8):1769-80
2. Accreditation Council for Graduate Medical Education. Common program requirements. Available at http://www.acgme.org/acWebsite/dutyHours/dh_dutyHoursCommonPR.pdf
3. Adams, J.A. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin* 1987; 101, 41-74.
4. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, Ramel S, Smith DC, Arvidsson D. Proficiency-based virtual reality training reduces the error rate for residents during their first 10 laparoscopic cholecystectomies *Am J Surg* 2007; 193: 797-804.
5. Akay M, March A. *Information Technologies in Medicine, Vol 1*. John Wiley and Sons Inc; 2001
6. Arnedt JT, Wilde GJ, Munt PW, MacLean AW. How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? *Accid Anal Prev* 2001;33:337-344
7. Asken MJ, Raham DC. Resident performance and sleep deprivation: a review. *J Med Educ* 1983;58:382-388.

8. Babineau B, Becker TJ, Gibbons G, Sentovich S, Hess D, Robertson R. The "Cost" of operative training for surgical residents, *Arch Surg* 2004;139: 366–369.
9. Bagwell CE. "Respectful Image": Revenge of the Barber Surgeon. *Ann Surg* 2005; 241(6): 872-878
10. Bailey RW, Imbembo AL, Zucker KA. Establishment of a laparoscopic cholecystectomy training program. *Am Surg* 1991; 57: 231-6.
11. Ball CG, Kirkpatrick AW, Feliciano DV, Reznick R, McSwain NE. Surgeons and astronauts: so close, yet so far apart. *Can J Surg.* 2008; 51(4): 247–250
12. Bencze JL. On-site science curriculum resources: constructivism. Available from: www.oise.utoronto.ca/lbencze/Constructivism.html.
13. Benner P. Using the Dreyfus Model of Skill Acquisition to Describe and Interpret Skill Acquisition and Clinical Judgment in Nursing Practice and Education. *Bulletin of Science, Technology & Society* 2004; 24 (3): 188-19
14. Bennett CL, Styker CJ. The learning curve for laparoscopic colorectal surgery. Preliminary results from a prospective analysis of 1194 lap-assisted colectomies. *Arch Surg* 1997; 132: 41-44
15. Berci G, Forde KA. History of endoscopy: what lessons have we learned from the past? *Surg Endosc* 2000;14(1): 5-15
16. Bernard GR. Prosection Demonstrations as Substitutes for the Conventional Human Gross Anatomy Laboratory *J Med Educ* 1972; 47:9
17. Bowles CJ, Leicester R, Romaya C, Swarbrick E, Williams CB, Epstein O. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow. *Gut.* 2004;53(2):277-83

18. Bridges M, Diamond DL, The financial impact of teaching surgical residents in the operating room, *Am J Surg* 1999; 177: 28–32.
19. Bruner J, *Toward a theory of instruction*, Cambridge (Mass), Belknap, 1966.
20. Brunton D. *Medicine Transformed – Health, Disease and Society in Europe 1800-1930*. Manchester, Manchester University Press, 2004
21. Bulstrode C, Hunt V. Selecting the best from the rest. *Surgeon*. 2003;1(6):328-31
22. Carlin AM, Gasevic E, Shepard AD. Effect of the 80-hour work week on resident operative experience in general surgery. *Am J Surg* 2007;193:326-30.
23. Carter KC, Carter BR. *Childbed Fever: A Scientific Biography of Ignaz Semmelweiss*. Westport, Connecticut: Greenwood Press; 1994
24. Cass OW, Freeman ML, Cohen J, Zuckerman G, Watkins J, Nord J, et al. Acquisition of competency in endoscopic skills (ACES) during training: A multicenter study. *Gastrointest Endosc* 1996;43:308.
25. Castellvi AO, Hollett LA, Minhajuddin A, Hogg DC, Tesfay ST, Scott DJ. Maintaining proficiency after fundamentals of laparoscopic surgery training: a 1-year analysis of skill retention for surgery residents. 2009;146(2):387-93
26. Cesanek P, Uchal M, Uranues S, Patruno J, Gogal C, Kimmel S, Bergamaschi R. Do simulator metrics correlate with content-valid outcome measures? *Sur Endosc* 2008; 22(10): 2178-83
27. Chaer RA, DeRubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D Morrissey NJ, Faries PL, McKinsey JF, Kent KC. Simulation improves resident performance in catheter-based intervention results of a randomised, controlled study. *Ann Surg* 2006; 244:343-52.

28. Chamberland C. Honor, Brotherhood, and the Corporate Ethos of London's Barber-Surgeons' Company, 1570–1640. *J Hist Med Allied Sci* 2009 64(3):300-332
29. Chmarra MK, Jansen FW, Grimbergen CA, Dankelman J Retracting and seeking movements during laparoscopic goal-oriented movements. Is the shortest path length optimal? *Surg Endosc*. 2008;22(4):943-9
30. Chung CS, Lehmann LS Informed consent and the process of cadaver donation *Arch Pathol Lab Med*. 2002;126(8):964-8
31. Chung RS. How much time do surgical residents need to learn operative surgery? *Am J Surg* 2005;190:351-3
32. Clinical Outcomes of Surgical Therapy Group. A Comparison of laparoscopically-assisted and open colectomy for colon cancer *N Eng J Med* 2004;350:2050-2059
33. Cohen J, Cohen SA, Vora KC, Xue X, Burdick JS, Bank S, Bini EJ, Bodenheimer H, Cerulli M, Gerdes H, Greenwald D, Gress F, Grosman I, Hawes R, Mullin G, Schnoll-Sussman F, Starpoli A, Stevens P, Tenner S, Villanueva G Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointest Endosc* 2006 ; 64 : 361-36
34. Committee on patient safety and quality improvement. ACOG committee opinion number 398, February 2008: fatigue and patient safety. *Obstetrics & Gynecology* 2008; 111: 471-4
35. Cope DH, Fenton-Lee D Assessment of laparoscopic psychomotor skills in interns using the MIST Virtual Reality Simulator: a prerequisite for those considering surgical training? *ANZ J Surg*. 2008;78(4):291-6.

36. Cotton PB, Connor P, McGee D, Jowell P, Nickl N, Schutz S, Leung J, Lee J, Libby E. Colonoscopy: practice variation among 69 hospital-based endoscopists *Gastrointest Endosc.* 2003;57(3):352-7
37. Cowie R. Measurement and modelling of perceived slant in surfaces represented by freely viewed line drawings. *Perception* 1988; 27(5): 505 – 540
38. Cuschieri A, Francis N, Crosby J, Hanna GB. What do master surgeons think of surgical competence and revalidation? *Am J Surg* 2001;182:110–6
39. Dankelman J, Wentink M et al. Does virtual reality training make sense in interventional radiology? Training skill-, rule- and knowledge-based behaviour. *Cardiovasc Internat Radiolol* 2004; 27:417-421
40. Darzi A, Mackay S. Assessment of surgical competence. *Qual Health Care.* 2001;10(Suppl 2):ii64–9
41. Darzi A, Munz Y. The impact of minimally invasive surgical techniques. *Annu Rev Med* 2004;55:223-37
42. Darzi A. Hand-assisted laparoscopic colorectal surgery. *Semin Laparosc Surg.* 2001;8(2):153-60.
43. Dashfield AK, Lambert AW, Campbell JK, Wilkins DC. Correlation between psychometric test scores and learning tying of surgical reef knots. *Ann R Coll Surg Engl.* 2001;83(2):139-43
44. Datta V, Chang A, Mackay S, Darzi A The relationship between motion analysis and surgical technical assessments. *Am J Surg.* 2002;184(1):70-3.
45. Datta V, Mackay S, Mandalia M, Darzi A The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model *J Am Coll Surg.* 2001 Nov;193(5):479-85..

46. Dawson D, Reid K. Fatigue, alcohol and performance impairment. *Nature* 1997;388:235-235
47. Department of Health and Children. Guidance on the implementation of the European Working Time Directive (2004) www.doh.ie
48. Department of Health.. Hospital Doctors: Training for the Future. The report of the Working Group on Specialist Medical Training (The Calman Report) London: DoH, 1993
49. Dimopoulos VG, Robinson JS 3rd, Fountas KN The pearls and pitfalls of skull trephination as described in the Hippocratic treatise "On Head Wounds". *J Hist Neurosci.* 2008;17(2):131-40.
50. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, Aptowicz C, Pack AI. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep* 1997;20:267-267
51. Dunn D, Nair R, Fowler S, McCloy R. Laparoscopic cholecystectomy in England and Wales: results of an audit by the Royal College of Surgeons in England. *Ann Royal Coll S Eng* 1994: 76; 269-75
52. Dutta S, Dunnington G, Blanchard MC, Spielman B, DaRosa D, Joehl RJ. "And doctor, no residents please!", *J Am Coll Surg* 2003; 197: 1012–1017.
53. Ekstrom R, French J, Harman H. Manual for Kit of Factor-referenced Cognitive Tests. Princeton,NJ:Educational Testing Service;,1976
54. Ende J. Feedback in clinical medical education. *JAMA* 1983; 250: 777-781.
55. Enochsson L, Isaksson B, Tour R, Kjellin A, Hedman L, Wredmark T, Tsai-Felländer L Visuospatial skills and computer game experience influence the performance of virtual endoscopy. *J Gastrointest Surg.* 2004;8(7):876-82

56. Entertainment trends in America - consumer tracking study at
www.NPD.com, updated March 2009 .
57. Eraut E, Chivers GE. Professions, Competence and Informal Learning.
Cheltenham, UK:Edward Elgar Publishing, 2005..
58. Ericsson E, Krampe RE, Tesch-Romer. The role of deliberate practice in the
acquisition of expert performance. *Psych Rev* 1993; 100: 363–406.
59. Evans AW, Aghabeigi B, Leeson R, O'Sullivan C, Eliahoo J. Are we really as
good as we think we are? *Ann R Coll Surg Engl.* 2002;84(1):54-6
60. Evans AW, Leeson RM, Petrie A. Reliability of peer and self-assessment
scores compared with trainers' scores following third molar surgery. *Med
Educ.*2007 ;41(9):866-72.
61. Feanny MA, Scott BG, Mattox KL, Hirschberg A. Impact of the 80-hour
work week on resident emergency operative experience. *Am J Surg*
2005;190:947-951
62. Ferlitsch A, Glauninger P, Gupper A, Schillinger M, Haefner M, Gangl A,
Schoefl R. Evaluation of a virtual endoscopy simulator for training in
gastrointestinal endoscopy. *Endoscopy* 2002;34:698 –702.
63. Fitts P, Posner MI. *Human Performance.* Monterey: Brooks/Cole; 1967
64. Flew T, Humphreys S.. *Games: Technology, Industry, Culture.* South
Melbourne: Oxford University Press; 2005
65. Franzese C.B., Stringer S.P. *The Evolution of Surgical Training: Perspectives
on Educational Models from the Past to the Future.* *Otolaryngol Clin North
Am* 2007; 40 (6):1227-1235
66. Gallagher AG, Cates CU. Approval of virtual reality training for carotid
stenting. *JAMA* 2004; 292: 3024-6

67. Gallagher AG, Cowie R, Crothers I, Jordan-Black JA, Satava RM. PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc*. 2003;17(9):1468-71
68. Gallagher AG, Leonard G, Traynor OJ. Role and feasibility of psychomotor and dexterity testing in selection for surgical training *ANZ J Surg*. 2009;79(3):108-13.
69. Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G, Smith CD, Satava RM. Virtual Reality Simulation for the Operating Room: Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training. *Ann Surg* 2005; 24:364-372
70. Gallagher AG, Smith CD. Human-factors lessons learned from the minimally invasive surgery revolution. *Semin Laparosc Surg* 2003; 10: 100-10
71. Gargiulo DA, Hyman NH, Hebert JC. Women in Surgery - Do We Really Understand the Deterrents? *Arch Surg*. 2006;141:405-408
72. Glaser AY, Hall CB, Uribe SJ, Fried MP. The effects of previously acquired skills on sinus surgery simulator performance. *Otolaryngol Head Neck Surg*. 2005;133(4):525-30.
73. Glaser AY, Hall CB, Uribe SJ, Fried MP. The effects of previously acquired skills on sinus surgery simulator performance. *Otolaryngol Head Neck Surg*. 2005;133(4):525-30.
74. Gold JP, Verrier ED, Mathisen DJ, Fullerton DA, Orringer MD. Successful implementation of a novel internet hybrid surgery curriculum: the early phase outcome of thoracic surgery prerequisite curriculum e-learning project. *Ann Surg* 2004; 240: 499–509

75. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy *Surg Endosc* 2003;17:1082-5
76. Greenfield, CL, Johnson AL, ArendsMW, Wroblewski AJ..Development of parenchymal abdominal organ models for use in teaching veterinary soft tissue surgery. *Vet. Surg* 1993; 22 (5): 357-362
77. Guillou PJ, Quirke P, Thorpe H, Walker J, Jayne DG, Smith AM, Heath RM, Brown JM; MRC CLASICC trial group. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer *The Lancet* 2005;365; 1718-1726
78. Hallmann-Mikolajczak A. Ebers Papyrus. The book of medical knowledge of the 16th century B.C. Egyptians. *Arch Hist Filoz Med.* 2004;67(1):5-14
79. Harewood GC, Lieberman DA. Colonoscopy practice patterns since introduction of medicare coverage for average-risk screening. *Clin Gastroenterol Hepatol.* 2004;2(1):72-7.
80. Harewood GC, Murray F, Winder S, Patchett S. Evaluation of formal feedback on endoscopic competence among trainees: the EFFECT trial. *Ir J Med Sci* 2008; 177: 253-256.
81. Harper JD, Kaiser S, Ebrahimi K, Lamberton GR, Hadley HR, Ruckle HC, Baldwin DD. Prior video game exposure does not enhance robotic surgical performance *J Endourol.* 2007;21(10):1207-10..
82. Hassan I, Gerdes B, Koller M, Dick B, Hellwig D, Rothmund M, Zielke A. Spatial perception predicts laparoscopic skills on virtual reality laparoscopy simulator. *Childs Nerv Syst.* 2007; 23(6):685-9

83. Hassan I, Weyers P, Maschuw K, Dick B, Gerdes B, Rothmund M, Zielke A.
Negative stress-coping strategies among novices in surgery correlate with poor virtual laparoscopic performance. *Brit J Surg* 2006; 93(12): 1554-9
84. Hassan I, You YN, Cima RR, Larson DW, Dozois EJ, Barnes SA. Pemberton J. Hand-assisted versus laparoscopic-assisted colorectal surgery: Practice patterns and clinical outcomes in a minimally-invasive colorectal practice. *Surg Endosc* 2008; 22(3): 739-743
85. Hassan I, You, YN, Cima R. R Larson, D. W, Dozois, E. J., Barnes, S. A.
86. Hatzinger M, Kwon ST, Langbein S, Kamp S, Häcker A, Alken P. Hans Christian Jacobaeus: Inventor of human laparoscopy and thoracoscopy. *J Endourol.* 2006;20(11):848-50
87. Hicks CM, Gonzales R, Morton M, Gibbons R, Wigton R, Anderson RJ.
Procedural experience and comfort level in internal medicine trainees. *J Gen Intern Med* 2000; 15: 716-722
88. Himmelmann L. From barber to surgeon- the process of professionalisation. *Sven Med Tidskr.* 2007;11(1):69-87
89. How much feedback is necessary for learning to suture? O' Connor A, Schwaitzberg SD, Cao CG. *Surg Endosc* 2008; 22:1614-9
90. Howard, DM. The Sanders Teacher. *Flight* 1910; 2(50): 1006-1007.
91. Howells NR, Auplish S, Hand GC, Gill HS, Carr AJ, Rees JL Retention of arthroscopic shoulder skills learned with use of a simulator. Demonstration of a learning curve and loss of performance level after a time delay. *J Bone Joint Surg Am* 2009;91(5):1207-13
92. <http://www.acgme.org>
93. http://www.barberscompany.org/historical_group.html

94. http://www.dohc.ie/issues/european_working_time_directive/joint_statement.pdf?direct=1
95. <http://www.fda.gov/ohrms/dockets/ac/04/transcripts/4033t1.htm>
96. http://www.nhmrc.gov.au/health_ethics/animal/surgeon.htm
97. <http://www.thejag.org.uk/>
98. Iqbal CW, Cullinane DC, Schiller HJ, Sawyer MD, Zietlow SP, Farley DR. Surgical management and outcomes of 165 colonoscopic perforations from a single institution. *Arch Surg.* 2008;143(7):701-6; discussion 706-7.
99. Jacobs M, Verdeja JC, Goldstein HS. Minimally invasive colon resection (laparoscopic colectomy). *Surg Laparosc Endosc.* 1991;1:144–150
100. Josefson D. New York's junior doctors work illegally long hours
BMJ 1998;316:1625
101. Kaufman HH, Wiegand RL, Tunick RH, Teaching surgeons to operate—principles of psychomotor skills training, *Acta Neurochirurgica* 1987; 87:1–7
102. Kazdin A. *Behavior Modification in Applied Settings*, 4th Edition. Pacific Grove, California: Brooks/Cole Publishing; 1989
103. Keirsey, D. *Please Understand Me II: Temperament, Character, Intelligence.* Del Mar, California: Prometheus Nemesis Book Company; 1998
104. Knowles M. *The adult learner: a neglected species* (4th ed). Houston, Texas: Gulf; 1980
105. Kogan JR, Holmboe ES, Hauer EK. Tools for direct observation and assessment of clinical skills of medical trainees. *JAMA.* 2009;302(12):1316-1326
106. Kohn LT, Corrigan JM. *To Err is Human: Building a Safer Health System.* Washington, DC. Institute of Medicine, 1999

- 107.Langer RM, Kahan BD. Alexis Carrell's legacy-visionary of vascular surgery and endorgan transplantation. *Transplantation Proceedings* 2002; 34: 1061-66
- 108.Lau WY, Leow CK, Li AK.. History of endoscopic and laparoscopic surgery. *World J Surg* 1997; 21(4): 444-53.
- 109.Lee TH, Meyer GS, Brennan TA. A middle ground on public accountability. *N Engl J Med*. 2004;350(23):2409-2412
- 110.Leung L, Becker CE. Sleep deprivation and house staff performance: update 1984-1991. *J Occup Med* 1992;34:1153-1160.
- 111.Lin PH, Bush RL, Peden EK, Zhou W, Guerrero M, Henao EA, Kougiass P, Mohiuddin I, Lumsden AB. Carotid artery stenting with neuroprotection: assessing the learning curve and treatment outcome. *Am J Surg*. 2005;190(6):850-7.
- 112.Lüning TH, Keemers-Gels ME, Barendregt WB, Tan AC, Rosman C. Colonoscopic perforations: a review of 30,366 patients. *Surg Endosc*. 2007;21(6):994-7
- 113.Lutz H-J, Michael R, Gahl B, Savolainen H. Is Carotid Endarterectomy a Trainee Operation? *World J Surg* 2009; 33(2): 242-245
- 114.MacRae H, Reznick R. Teaching surgical skills – changes in the wind. *Med Ed* 2006; 355: 2664-9
- 115.Madan AK, Frantzides CT, Park WC, Tebbit CL, Kumari NV, O'Leary PJ Predicting baseline laparoscopic surgery skills. *Surg Endosc* 2005; 19(1):101-4.
- 116.Madan AK, Harper JL, Frantzides CT, Tichansky DS. Nonsurgical skills do not predict baseline scores in inanimate box or virtual-reality trainers *Surg Endosc*. 2008;22(7):1686-9.

117. Magnier LM. A history of medicine. New York: Marcel Dekker Inc; 1992
118. Maharaj CH, Costello J, Higgins BD, Harte BH, Laffey JG. Retention of tracheal intubation skills by novice personnel: a comparison of the Airtraq and Macintosh laryngoscopes. *Anaesthesia*. 2007;62(3):272-8.
119. Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surg Endosc* 2004; 8:1224-30
120. Marcello PW, Fleshman JW, Milsom JW, Read TE, Arnell T, Birnbaum EH, Feingold DL, Lee SW, Mutch MG, Sonoda T, Yan Y. Hand-assisted laparoscopic vs. laparoscopic colorectal surgery: a multicenter, prospective, randomized trial. *Dis Colon Rectum* 2008; 51(6): 818-326
121. Marteniuk, R. Information Processing in Motor Skills. New York: Holt, Rinehart & Winston; 1976
122. Mas J-L, Chatellier G, Beyssen B, Branchereau A, Moulin T, Becquemin JP, Larrue V, Lièvre M, Leys D, Bonneville JF, Watelet J, Pruvo JP, Albucher JF, Viguier A, Piquet P, Garnier P, Viader F, Touzé E, Giroud M, Hosseini H, Pillet JC, Favrole P, Neau JP, Ducrocq X; EVA-3S Investigators
Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis. *N Engl J Med* 2006;355:1660-1671.
123. Maschuw K, Osei-Agyemang T, Weyers P, Danila R. Bin Dayne K, Rothmund M, Hassan I. The Impact of Self-Belief on Laparoscopic Performance of Novices and Experienced Surgeons. *World J Surg* 2008; 32:1911–1916

124. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effects of bench model fidelity on endourological skills: a randomised controlled study, *J Urol* 167; 243–247.
125. McCallum JE. *Military Medicine: From Ancient Times to the 21st Century*. California: ABC-Clio; 2008
126. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007; 7: 30.
127. McClusky DA 3rd, Ritter EM, Lederman AB, Gallagher AG, Smith CD
Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *Am Surg*. 2005;71(1):13-20
128. McClusky DA, Smith CD. Design and Development of a Surgical Skills Simulation Curriculum. *World J Surg* 2008; 32: 171-181
129. McLachlan JC, Bligh J, Bradley P, Searle J. Teaching anatomy without cadavers. *Med Ed* 2004; 28: 418-24
130. Mery CM, Cooke DT et al. The Road to Innovation. *Bulletin Am C S* 2007; 92(7): 20-3
131. Mery CM, Cooke DT, Chandra V, Shafi BM, Tavakkolizadeh A, Varghese TK. The road to innovation: emerging technologies in surgery. *Bull Am Coll Surg* 2007; 92: 20-33
132. Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care*. 2009;18(2):104-8.

133. Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R
Teaching surgical skills: what kind of practice makes perfect?: a randomised, controlled trial. *Ann Surg.* 2006;244(3):400-9
134. Muhe, E. Long-term follow-up after laparoscopic cholecystectomy.
Endoscopy 1992; 24(9): 754-8
135. Neary PC, Boyle E, Delaney CP, Senagore AJ, Keane FB, Gallagher AG.
Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc* 2008; 22: 2301-9.
136. Neequaye SK, Aggarwal R, Van Herzeele I, Darzi A, Cheshire NJ.
Endovascular skills training and assessment. *J Vasc Surg* 2007; 46(5):1055-64.
137. Nomura T, Miyashita M, Shrestha S, Makino H, Nakamura Y, Aso R, Yoshimura A, Shimura T, Akira S, Tajiri T. Can interview prior to laparoscopic simulator training predict a trainee's skills? *J Surg Educ.* 2008;65(5):335-9.
138. Nordmann AJ, Woo K, Parkes R, Logan AG. Balloon angioplasty or medical therapy for hypertensive patients with atherosclerotic renal artery stenosis? A meta-analysis of randomized controlled trials. *Am. J. Med* 2003 114 (1): 44–50
139. Novielli K, Hojat M, Park PK, Gonnella JS, Veloski JJ. Change of interest in surgery during medical school: a comparison of men and women. *Acad Med.* 2001;76 (10 suppl):S58-S61.
140. Nutton V. The Chronology of Galen's Early Career. *Class Q* 1973;23(1):158-71

141. O' Connor A, Schwaitzberg SD, Cao CG. How much feedback is necessary for learning to suture? *Surg Endosc* 2008; 22:1614-9
142. O'Shea J. Becoming a Surgeon in the 20th Century: Parallels to the present. *J Surg Education* 2008; 65: 236-40
143. O'Malley, C., *Andreas Vesalius of Brussels, 1514-1564*. Berkeley: University of California Press: 1964
144. Owens JA. Sleep loss and fatigue in medical training. *Curr Opin Pulm Med* 2001;7:411-418.
145. Ozkara, H, Watson LR. Laparoscopic surgery in urology. *Int Urol Nephrol* 1992; 24(5): 461-4.
146. Pandey VA, Wolfe JH, Black SA, Cairns M, Liapis CD, Bergqvist D; European Board of Vascular Surgery Self-assessment of technical skill in surgery: the need for expert feedback. *Ann R Coll Surg Engl*. 2008;90(4):286-90.
147. Park J, MacRae H, Musselman LJ, Rossos P, Hamstra SJ, Wolman S, Reznick RK. Randomized controlled trial of virtual reality simulator training: transfer to live patients. *Am J Surg* 2007;194(2):205-11.
148. Pellen M, Horgan L, Roger Barton J, Attwood S. Laparoscopic surgical skills assessment: can simulators replace experts? *World J Surg*. 2009 33(3):440-447
149. Pellen MG, Horgan LF, Barton JR, Attwood SE. Construct validity of the ProMIS laparoscopic simulator. *Surg Endosc* 2009;23(1):130-9.
150. Pemberton, J. Hand-assisted versus laparoscopic-assisted colorectal surgery: Practice patterns and clinical outcomes in a minimally-invasive colorectal practice *Surg Endosc* 2007;22(3):739-43

151. Persaud TVN. The Early History of Human Anatomy: From Antiquity to the Beginning of the Modern Era. Springfield: Thomas Books; 1984
152. Polk HC Jr. The declining interest in surgical careers, the primary care mirage, and concerns about contemporary undergraduate surgical education. *Am J Surg*. 1999;178:177-179.
153. Polychronidis A, Laftsidis P, Bounovas A, Simopoulos C. Twenty years of laparoscopic cholecystectomy: Philippe Mouret--March 17, 1987. *JLS*. 2008 ;12(1):109-11
154. Pories WJ. Some reflections on Halsted and Residency Training. *Current Surg* 1999;
155. Price J, Campbell C, Sells J, Weller D, Campbell H, Kenicer M, Dunlop M.J Impact of UK Colorectal Cancer Screening Pilot on hospital diagnostic services. *J Public Health (Oxf)*. 2005;27(3):246-53.
156. Principles of Training in Gastrointestinal Endoscopy. *Gastrointest Endosc* 1999; 49: 845-853
157. Provision of endoscopy-related services in district general hospitals. Working Party of the British Society of Gastroenterology Endoscopy Committee. London: British Society of Gastroenterology Endoscopy Committee, working party report, 2001.
<http://www.bsg.org.uk/clinicalguidelines/endoscopy/index.html>
158. Ravi S, Sidhu RS, Park J, Brydges R, MacRae H, Dubrowski A. Laboratory-based vascular anastomosis training: A randomised controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. *J Vasc Surg* 2007 45(2) 342-349

159. Recommendations for training in gastrointestinal endoscopy 1999. The Joint Advisory Group on Gastrointestinal Endoscopy: London, JCHMT, 1999
160. researchers and practitioners. Boston: Kluwer Academic Publishers; 2001
161. Rex DK, Bond JH, Winawer S, Levin TR, Burt RW, Johnson DA, Kirk LM, Litlin S, Lieberman DA, Wayne JD, Church J, Marshall JB, Riddell RH; U.S. Multi-Society Task Force on Colorectal Cancer Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U.S. Multi-Society Task Force on Colorectal Cancer. *Am J Gastroenterol.* 2002 ;97(6):1296-308
162. Rex DK, Johnson DA, Lieberman DA, Burt RW, Sonnenberg A. Colorectal cancer prevention 2000: screening recommendations of the American College of Gastroenterology. *Am J Gastroenterol* 2000; 95 (4): 868–77.
163. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W.. Testing technical skills via an innovative bench station examination. *Am J Surg* 1997;173:226-3
164. Ritter EM, Kindelan TW, Michael C, Pimentel EA, Bowyer MW. Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). *Surg Endosc.* 2007; 21(8):1441-5.
165. Ritter EM, McClusky DA 3rd, Gallagher AG, Enochsson L, Smith CD. Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. *Am J Surg.* 2006;192(3):379-84.
166. Roberts JS, Radany MH, Nash DB. Privilege delineation in a demanding new environment. *Ann Intern Med.* 1988; 108:880-6

167. Rosenberg BH, Landsittel D, Averch TD Can video games be used to predict or improve laparoscopic skills? *J Endourol.* 2005; 19(3):372-6.
168. Rosser JC, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. *Arch Surg* 2007;142(2):181-6
169. Rowe G, Wright G. Expert opinions in forecasting: role of the Delphi technique. In: Armstrong. *Principles of forecasting: A handbook of*
170. Samkoff JS, Jacques CH. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Acad Med* 1991;66:687-693.
171. Satava RM. Virtual reality surgical simulator; the first steps. *Surg Endosc* 1993; 7: 203-205.
172. Savion-Lemeux T, Penhume VP. The effects of practice and delay on motor skill learning and retention. *Experimental brain research* 2005;161(4):423-31
173. Schijven M, Jakimowicz J, Carter F. How to select aspirant laparoscopic surgical trainees: Establishing concurrent validity comparing Xitact LS500 index performance scores with standardized psychomotor Aptitude Test Battery scores. *J Surg Research* 2004; 121 (1): 112-119.
174. Schijven MP, Jakimowicz JJ, Carter FJ. How to select aspirant laparoscopic surgical trainees: establishing concurrent validity comparing Xitact LS500 index performance scores with standardized psychomotor aptitude test battery scores. *J Surg Res.* 2004;121(1):112-9.
175. Schollmeyer T, Soyinka AS, Schollmeyer M, Meinhold-Heerlein I. Georg Kelling (1866-1945): the root of modern day minimal invasive surgery. A forgotten legend? *Arch Gynecol Obstet.* 2007;276(5):505-9.

176. Schueneman AL, Pickleman J, Hesslein R, Freeark RJ. Neuropsychologic predictors of operative skill among general surgery residents. *Surgery*. 1984;96:288–95
177. Sedlack RE, Kolars JC. Colonoscopy curriculum development and performance-based assessment criteria on a computer-based endoscopy simulator. *Acad Med*. 2002 ;77(7):750-1.
178. Seeff LC, Richards TB, Shapiro JA, Nadel MR, Manninen DL, Given LS, Dong FB, Wings LD, McKenna MT. How many endoscopies are performed for colorectal cancer screening? Results from CDC's survey of endoscopic capacity. *Gastroenterology*. 2004;127(6):1670-7.
179. Semm, K. Endoscopic appendectomy. *Endoscopy* 1983; 15(2): 59-64
180. Sewell C, Morris D, Blevins NH, Dutta S, Agrawal S, Barbagli F, Salisbury K. Providing metrics and performance feedback in a surgical simulator. *Comput Aided Surg*. 2008 ;13(2):63-81
181. Seymour N, Gallagher A, Sanziana R, O' Brien MK, Bansal V, Andersen DK, Satava R. Virtual reality training improves operating room performance: results of a randomised, double-blinded study. *Ann Surg* 2002; 236: 458-464.
182. Shah HA, Paszat LF, Saskin R, Stukel TA, Rabeneck L. Factors associated with incomplete colonoscopy: a population-based study. *Gastroenterology*. 2007 ;132(7):2297-303.
183. Sickle K, Gallagher AG, Smith CD. The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. *Surg Endosc* 2007; 21: 220-224
184. Singer, RN. *Motor Learning and Human Performance* (2nd Ed.). New York: Macmillan;1975

185. Sinha P, Hogle NJ, Fowler DL. Do the laparoscopic skills of trainees deteriorate over time? *Surg Endosc.* 2008;22(9):2018-25.
186. Spritz N. Oversight of physicians' conduct by state licensing agencies. Lessons from New York's Libby Zion case. *Ann Intern Med* 1991;115(3):219-2
187. Stefanidis D, Korndorffer JR Jr, Black FW, Dunne JB, Sierra R, Touchard CL, Rice DA, Markert RJ, Kastl PR, Scott DJ . Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. *Surgery* 2006;140(2):252-62.
188. Stefanidis D, Korndorffer JR Jr, Heniford BT, Scott DJ Limited feedback and video tutorials optimise learning and resource utilisation during laparoscopic simulator training., *Surgery* 2007;142:202-6
189. Stefanidis D, Korndorffer JR Jr, Sierra R, Touchard C, Dunne JB, Scott DJ Skill retention following proficiency-based laparoscopic simulator training. *Surgery.* 2005;138(2):165-70
190. Sternhouse L. An introduction to curriculum research and development. London: Heineman; 1975
191. Strongwater AM Transition to the eighty-hour resident work schedule *J Bone Joint Surg Am* 2003; 85(6): 1170-2
192. Sutherland, I. A head-mounted three-dimensional display. In: Proceedings of the Fall Joint Computer Conference. AFIPS Conference Proceedings, vol. 33. AFIPS, Arlington, VA., 757- 764
193. Tansley P, Kakar S, Withey S, Butler P. Visuospatial and Technical Ability in the Selection and Assessment of Higher Surgical Trainees in the London Deanery. *Ann R Coll Sur Engl* 2007; 89(6): 591-595

194. Tashiro Y, Miura H, Nakanishi Y, Okazaki K, Iwamoto Y. Evaluation of skills in arthroscopic training based on trajectory and force data. *Clin Orthop Relat Res.* 2009;467(2):546-52.
195. Tekis P, Senagore A. Evaluation of the Learning Curve in Laparoscopic Colorectal Surgery: Comparison of Right-Sided and Left-Sided Resections. *Ann Surg* 2005; 242 : 83-91
196. The colon cancer laparoscopic or open resection study group. Laparoscopic surgery for colon cancer: short-term outcomes of a randomized trial *Lancet Oncol* 2005;6:470-484
197. The Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies *N Eng J Med* 1991; 324: 1073-78
198. The Southern Surgeons Club. The learning curve for laparoscopic cholecystectomy. *Am J Surg* 1995; 170: 55-59
199. Tsai CL, Heinrichs WL. Acquisition of Eye-hand Coordination Skills for Videoendoscopic Surgery. *Am Assoc Gynecol Laparosc* 1994; 1(4): 4-9
200. Tsang PS, Vidulich M. Principles and practice of aviation psychology. Lawrence Erlbaum Associates: New York; 2003
201. Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *J Am Board Fam Pract* 1998;11(6):426-33.
202. Turner LWF. Teaching Flying. *Flight* 1913; 5(12): 326-328
203. Uribe JI, Ralph WM Jr, Glaser AY, Fried MP. Learning curves, acquisition, and retention of skills trained with the endoscopic sinus surgery simulator. *Am J Rhinol.* 2004;18(2):87-92.

204. Van Dongen HPA, Dinges DF. Circadian rhythms in fatigue, alertness, and performance. In: Kryger MH, Roth T, Dement WC, eds. Principles and practice of sleep medicine (3rd ed.) Philadelphia: W.B. Saunders; 2000
205. Van Herzele I, Aggarwal R, Neequaye S, Darzi A, Vermassen F, Cheshire NJ. Cognitive training improves clinically relevant outcomes during simulated endovascular procedures. *J Vasc Surg.* 2008;48(5):1223-30
206. Van Herzele I, Aggarwal R, Neequaye S, Hamady M, Cleveland T, Darzi A, Cheshire N, Gaines P. Experienced endovascular interventionalists objectively improve their skills by attending carotid artery stent training courses. *Eur J Vasc Endovasc Surg.* 2008;35(5):541-50.
207. Van Sickle KR, Ritter EM, McClusky DA 3rd, Lederman A, Baghai M, Gallagher AG, Smith CD. Attempted establishment of proficiency levels for laparoscopic performance on a national scale using simulation: the results from the 2004 SAGES Minimally Invasive Surgical Trainer-Virtual Reality (MIST-VR) learning center study. *Surg Endosc* 2007;21:5-11
208. Velmahos GC, Toutouzas KG, Sillin LF, Chan L, Clark RE, Theodorou D. Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory, *Am J Surg* 2004; 187:,114–119
209. Vilos GA, Ternamian A, Dempster J, Laberge PY Laparoscopic entry: a review of techniques, technologies, and complications. *J Obstet Gynaecol Can.* 2007;29(5):433-65
210. Wagner OJ, Hagen M, Morel P, Inan I, Candinas D, Vorburger SA Who should do NOTES? Initial endoscopic performance of laparoscopic surgeons compared to gastroenterologists and untrained individuals *J Gastro Intest Surg* 2008; 12(10): 1724-9

211. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA*. 2005;293(10):1239-1244.
212. Whitehouse A, Hassell A, Bullock A, Wood L, Wall D. 360 degree assessment (multisource feedback) of UK trainee doctors: field testing of team assessment of behaviours (TAB) *Med Teach*. 2007 Mar;29(2-3):171-6
213. Wigton RS. Measuring procedural skills (editorial). *Ann Intern Med* 1996; 125: 1003-4.
214. Winstein CJ, Schmidt RA. *J Exp Psychol Learn Mem Cogn* 1990;16:677-91
215. Wiwanitkit V, Agthong S. The rate of anti-HIV seropositivity among donated cadavers: experience in a cadaver donation centre. *Viral Immunol*. 2002;15(4):659-60
216. www.nlm.nih.gov/research/visible/visible_human.html
217. Yadav JS, Wholey MH, Kuntz RE, Fayad P, Katzen BT, Mishkel GJ, Bajwa TK, Whitlow P, Strickman NE, Jaff MR, Popma JJ, Snead DB, Cutlip DE, Firth BG, Ouriel K; Stenting and Angioplasty with Protection in Patients at High Risk for Endarterectomy Investigators. Protected carotid-artery stenting versus endarterectomy in high-risk patients. *N Engl J Med* 2004;351:1493-1501.
218. Zucker, KA. *Surgical Laparoscopy*. Philadelphia: Lippincott Williams and Wilkins; 2001

Appendices

Appendices

Appendix Chapter 3

Appendix 3.1

POST- TRAINING FEEDBACK

Subject code: _____

Date: _____

Agree strongly	Agree	Neutral	Disagree	Disagree strongly	
5	4	3	2	1	
Simulator-based training is useful in surgical training	5	4	3	2	1
Training on a simulator helps improves real surgical performance	5	4	3	2	1
Simulator training should be obligatory for all trainees	5	4	3	2	1
In your opinion, will simulator training become obligatory for all trainees	5	4	3	2	1
Simulator-based training will improve patient safety	5	4	3	2	1

Was today beneficial for you?	5	4	3	2	1
Do you think this training will improve your operative skills?	5	4	3	2	1
Would you feel more confident about performing laparoscopic colorectal surgery after today?	5	4	3	2	1
Would you feel more confident about performing laparoscopic surgery in general after today?	5	4	3	2	1
Would you be happy to attend a similar course again?	5	4	3	2	1

Any other feedback on any aspect of the simulator training: _____

Appendix 3.2 - Tray error definitions

1. Incomplete division of the inferior mesenteric artery
2. Incomplete division of the inferior mesenteric vein
3. Mesenteric injury
4. Inadequate exposure of the left ureter
5. Inadequate division of the sigmoid mesentery
6. Inadequate lateral mobilisation of the colon
7. Inadequate mobilisation of the splenic flexure
8. Inadequate division of the mesorectum
9. Inadequate rectal transection
10. Inadequate anastomotic alignment
11. Excessive anastomotic tension
12. Asymmetrical anastomosis
13. Organ injury
14. Inadequate specimen resection

Appendix 3.3

Proficiency levels for the HALC on the ProMIS simulator

Instrument Path Length	3,862 mm
Instrument Smoothness	375 counts
Tray error score	2 errors

Appendix 3.4

Subject self assessment form

Subject code: _____

Procedure code: _____

How many errors did you make? _____

Name errors (where possible) :

Please rate your performance on the following scale:

Respect for tissue	1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments	2	3 Careful handling of tissues but occasionally caused inadvertent damage.	4	5 Consistently handled tissues appropriately with minimal damage.
Time and motion	1 Many unnecessary moves.	2	3 Efficient time/motion but some unnecessary moves.	4	5 Economy of movement and maximum efficiency.
Instrument handling	1 Repeatedly makes tentative or awkward moves with instruments	2	3 Competent use of instruments although occasionally appeared stiff or awkward.	4	5 Fluid moves with instruments and no awkwardness.
Knowledge of instruments	1 Frequently asked for the wrong instrument or used an inappropriate instrument.	2	3 Knew the names of most instruments and used appropriate instrument for the task.	4	5 Obviously familiar with the instruments required and their names.
Use of assistants	1 Consistently placed assistants poorly or failed to use assistants.	2	3 Good use of assistants most of the time.	4	5 Strategically used assistants to the best advantage at all times.
Flow of operation and forward planning	1 Frequently stopped operating or needed to discuss next move.	2	3 Demonstrated ability for forward planning with steady progression of operative procedure.	4	5 Obviously planned course of operation with effortless flow from one move to the next.
Knowledge of specific procedure	1 Deficient knowledge. Needed specific instruction at most operative steps.	2	3 Knew all important aspects of the operation.	4	5 Demonstrated familiarity with all aspects of the operation.

How would you rate your performance overall? (1 – 5 scale, where 1 is poor, 2 is sub-standard, 3 is average, 4 is good and 5 is excellent) _____

Appendix Chapter 4

Appendix 4.1 – Standardised advice given to subjects

Advice Given

General

Fluoro use (stop when changing instruments, etc)

Contrast use (roadmaps)

Instruments

Stabilise instruments while threading others over/under

Watch distal points – aorta, renal

Angio

Obtain full view RS

Withdraw GW to femoral if necc to avoid buckling

Roadmap

Cannulation

Angle catheter laterally

Withdraw slowly

Watch for “give”

Insert GW over lesion but not into parenchyma

Stabilise guidewire

Guide catheter

Angle of sheath to avoid pulling guidewire

Renal ostium, avoid lesion

Angioplasty

Choose appropriate balloon – diameter/length

Centre balloon over lesion

Deflate before withdrawing

Stenting

Stabilise GW

Choose appropriate stent – diameter/length

Centre stent over lesion

Deflate before w/d

VIST Renal Artery Procedure

Candidate instruction form

- Insert 0.035 guidewire to appropriate level
- Insert diagnostic catheter
- Withdraw guidewire
- Perform angiogram to assess renal vasculature. (Ask for roadmap to be taken if required)
- Cannulate renal artery with diagnostic catheter (ie judkins)
- Repeat angiogram if necessary
- Reintroduce 0.035 guidewire, insert into renal artery
- Exchange diagnostic catheter for guide catheter, insert to renal ostium
- Select balloon, insert balloon over guidewire, position over stenosis
- Inflate balloon to 12 atm, deflate and withdraw
- Select balloon expandable stent, deploy, withdraw
- Perform completion angiogram

Appendix 4.3 – subject self assessment sheet

VIST study self assessment sheet.

Subject code: _____

Procedure 1 2 3 4 5 6

		Very poor		average		Very good
1	How well did you know the steps of the procedure?	1	2	3	4	5
2	How would you rate your anatomical knowledge during the case?	1	2	3	4	5
3	How was your knowledge of the instruments?	1	2	3	4	5
4	How would you rate your ease of instrument use?	1	2	3	4	5
5	How would you rate yourself regarding respect for vessels?	1	2	3	4	5
6	How safe was the procedure?	1	2	3	4	5
7	How was your efficiency of fluoroscopy use?	1	2	3	4	5
8	How well did you recognise problems that occurred, if any?	1	2	3	4	5
9	How efficient were you with regard to length of procedure?	1	2	3	4	5
10	How would you rate your overall performance	1	2	3	4	5

Did you make any errors? Yes/no

Please list any errors made

Did you use the correct instrument for each step? Yes/no

Do you think you endangered patient safety? Yes/no

Will you do anything differently on your next case? yes/no

Details: _____

Appendix 4.4 – part of video assessment sheet

	steps	errors		sc
--	-------	--------	--	----

1	Advance guidewire into aorta	Advance without following		
		Advance to wrong position >T12, <L3		
2	Introduce pigtail	wrong instrument		
		fail to visualise GW tip while inserting		
		fail to stabilise GW		
		advance too far (>t12), not enough (<l3)		
		fail to pull back slightly on pigtail		
3	Obtain renal angiogram	insufficient view of RA, need to repeat		
		fail to repeat if insufficient		
4	Remove pigtail	GW curls/buckles in aorta		
		move too fast		
		fail to visualise		
5	Re insert GW	Advance without following		
		Advance to wrong position >T12, <L3		
6	Insert diagnostic catheter to L1	wrong instrument		
		fail to visualise GW tip while inserting		
		fail to stabilise GW		
		advance too far (>t12), not enough (<l3)		
		fail to pull back slightly on catheter		
7	Cannulate renal artery	fail to rotate to left		
		fail to maintain left position		
		fail to withdraw slowly		
		fail to recognise if tip engages with RA		
		cannulate wrong vessel		
		cannulate but pull catheter out		
		need to repeat, (have to pass catheter prox to RA again)		
		fail to obtain angiogram		
		insufficient, ie stenosis not visualised		
		fail to repeat if insufficient		
8	Insert guidewire into RA, exchange diagnostic catheter for guide catheter	insert GW too far into RA		
		fail to stabilise GW		
		pull GW out of RA when removing catheter		
		guide catheter angled to right		
		pull GW out when inserting guide catheter		
		need to re peat steps 5-9		
		G catheter not engaged with RA ostium		
		insert guide catheter past lesion		
9	Insert balloon	failure to use roadmap to guide size		
		inappropriate size		
		insert GW too far		
		pull guide catheter out of RA		
		no use roadmap		
		fail to centre balloon over lesion		
10	Angioplasty	inflate inside guide catheter		

11	Exchange balloon for stent				
12	Deploy stent				
13	Completion angiogram				
14	Removal of instruments				

Procedure finished comments

variable	0	1	2	3	4
Respect for tissue	Often used unnecessary force on tissue or lesion, potential for tissue damage		Careful handling of tissue/lesion, but occasional potential for inadvertent damage		Consistently handled tissue/lesion appropriately with minimal damage

variable	0	1	2	3	4
Handling of wire/ catheters	Many unnecessary, awkward or inefficient movements, difficulty controlling instruments		Generally competent use of instruments few unnecessary moves, occasionally awkward		Demonstrates consist ease and appropriateness of use of instrument, fluid movements

Variable	0	1	2	3	4
Response to difficulties	Unaware of cause when encountering difficulty, repeats same mistakes		Recognises difficulties, attempts to resolve		Recognises and takes appropriate action when encounters difficulty

Variable	0	1	2	3	4
Knowledge and flow of procedure	Frequently stopped case, unsure of next move		Reasonably knowledgeable, occasional hesitation		Familiar with all aspects of procedure, easy flow

Variable	0	1	2	3	4
General safety of procedure (instruments, contrast, lesion etc)	<i>Generally poor, frequently suboptimal, risk of morbidity</i>		<i>Reasonably good, aware of consequences of actions</i>		<i>Safe performance consistently, low risk of patient morbidity</i>

variable	0	1	2	3	4
Overall performance	Very poor		average		excellent

Appendix 4.5 – sample procedural report generated by the simulator

Procedure report

Date: 2009-02-26, 01:16

User: xxxx

Module: Generic Renal Intervention

Case: 2

General Information

Total time 00: 18: 06

Total amount of contrast used 83.9 cc

Total fluoroscope time 00: 14: 00

Per Lesion Report

Lesion: Left lesion

Dilatation Catheter, 5, 15

Placement Accuracy proximal (centre lesion - centre balloon) 5.2 mm

Diameter of Balloon (inflated) 3.3 mm

Inflation Time 7.4 s

Diameter stenosed (before dilatation) 75 %

Residual stenosis 18 %

Reference lumen diameter 4.0 mm

% of lesion covered with Balloon 100 %

Balloon - vessel ratio 0.8

Max pressure reached 12.8 atm

Lesion length 12.4 mm

Balloon Expandable Stent, 5, 15

Placement Accuracy proximal (centre lesion - centre stent) 4.5 mm

Diameter of Stent (inflated) 4.6 mm

Inflation Time 6.3 s

Diameter stenosed (before dilatation) 75 %

Residual stenosis 0 %

Reference lumen diameter 4.0 mm

% of lesion covered with Stent 100 %

Stent - vessel ratio 1.2

Max pressure reached 12.5 atm

Lesion length 12.4 mm

Handling Errors

Catheter scraping against vessel wall 0

Catheter moving without support of wire 2

Selective catheter scraping against vessel wall 3

Selective catheter moving without support of wire 1

0.014 wire hit end of vessel 0

0.014 wire entered suboptimal vessel 0

Guide wire in small vessel 0

Guide wire entered suboptimal vessel 4

Catheter entered suboptimal vessel 0

Report Details

00: 10: 41 Balloon applied on Left lesion using:

Dilatation Catheter, 5, 15

00: 15: 26 Stent applied on Left lesion using:

Balloon Expandable Stent, 5, 15

00: 18: 06 Procedure finished

Used Tools

Standard Guide Wire guide wire, diameter 0.035, shape Standard Tip

Diagnostic Catheter, Pigtail

Diagnostic Catheter, Judkins Right 4

Guide Catheter, Multipurpose A

Appendix Chapter 7

Appendix 7.1 Subject information sheet

Subject Information Sheet

The Benefits of a Proficiency-Based Virtual Reality Simulation Curriculum in the Training of Future Providers of Colonoscopy in Ireland: Skills Transfer from the Virtual to the Real Environment.

What is the study about?

The study is about training in colonoscopy, and how to improve it. The aim of this research project is to demonstrate that medical and surgical trainees who undergo colonoscopy training in a structured, proficiency-based progression curriculum which incorporates the use of virtual reality (VR) simulation will transfer these skills to the real, clinical environment. We plan to achieve this by training a group of trainees in colonoscopy, and comparing their performance during their first five cases to the performance of a matched group of trainees who have been trained in the traditional manner, in other words the normal training which occurs currently in hospitals around Ireland.

Secondary aims of the study include comparing the results of the simulator generated performance metrics with the results generated by the checklist assessment forms currently used at the RCSI for the selection of higher surgical trainee and defining a national proficiency level for colonoscopy on the GI Mentor simulator. In addition, if the VR curriculum is shown to produce superior results, the study findings may encourage the development of a standardised training programme in colonoscopy.

Your participation in this trial is voluntary.

What happens if I agree to take part?

If you agree to take part in the study, you first be asked to sign a consent form, Then you will be randomised to either the simulator curriculum (SC) group or the conventional curriculum (CC) group. All trainees will then undergo a series of tests of perception, visuospatial ability and psychomotor coordination to assess their fundamental abilities. They will also access a didactic teaching programme which will partly consist of web-based teaching materials and partly a directed lecture. They will have to take an MCQ following this and score 100%. Then all subjects will perform a “pre-test” on the simulator which involves performing one simulated colonoscopy.

Subjects in the SC Group will then take part in a structured training curriculum, which will involve performing simulated cases on the GI Mentor simulator until they have reached a predefined level of proficiency.

Subjects in the CC group will continue to receive the standard training offered at their hospital. This may include observing, assisting and performing parts of a colonoscopy.

Subjects in both groups will have had some exposure to colonoscopy at their hospitals before and during the trial, but no subject can have performed a complete procedure before their trial cases.

After the training, all subjects will proceed to perform five colonoscopies at their hospitals under the supervision of an experienced colonoscopist. They will be assessed by their supervisor, and an independent evaluator using a standardised assessment form. Patients who satisfy specific criteria, ie those whose colonoscopies are presumed to be suitable for a trainee will be selected for the study. Results from the two groups, ie SC group and CC group will be analysed and compared

Implications

Once enrolled in the trial, any data pertaining to you is handled in a confidential manner. Results from the fundamental abilities tests, MCQ, simulator performance , and performance in the endoscopy suite itself will not be released to or shared with any surgical or medical training bodies and will have no effect on your career. Any cases you perform as part of the trial can be entered in your logbook, but our assessment of your performance will only be used in the context of the trial itself. As is the case with regard to any procedures you carry out in the hospital, you are covered by the Hospital Clinical Indemnity Scheme for these cases. Taking part in the trial will give you access to a didactic teaching programme and the chance to perform five colonoscopy cases with close supervision. Although your rotation should expose you to these cases anyway, with pressure of time in the endoscopy suite, it can be difficult to perform colonoscopies as a trainee, so there are obvious benefits to participating in the study. In addition, those in the SC group will also have the chance to take part in a structured simulator training programme, and train to a level of proficiency. This will be a valuable learning opportunity

Contact Details

If you have any questions about the study, you can talk to the consultant who is coordinating the study in your hospital, or to a member of the study management team at RCSI.

RCSI contact - xxxx

Supervisor – xxxx

Appendix 7.2 Subject Consent form

FORM OF CONSENT BY A SUBJECT VOLUNTEERING TO TAKE PART IN RESEARCH ASSOCIATED WITH SURGICAL TRAINING AND

CONSENT BY THE SUBJECT

I _____ (full name)

of _____ (address)

hereby fully and freely agree to take part in a study entitled

“The Benefits of a Proficiency-Based Virtual Reality Simulation Curriculum in the Training of Future Providers of Colonoscopy in Ireland: Skills Transfer from the Virtual to the Real Environment.”

I have read and understand the subject information sheet.

I understand that any data collected during the trial which pertains to me will be treated in a confidential manner.

I understand that this study is designed to promote scientific knowledge, develop surgical training and improve patient care.

I agree to fulfil any training requirements to the best of my ability, and undertake to maintain a detailed logbook of all operating activity for the duration of my involvement in the trial. This may be inspected by investigators/researchers connected with the trial

I understand that I may withdraw my consent at any stage in the study without having to provide an explanation. I acknowledge the purpose of the study. The nature and purpose of any procedure has been described and explained to me by:

_____ (name of researcher)

and I have discussed these matters with him/her to my satisfaction.

Signed: _____

Print name: _____

Date: _____

DECLARATION BY THE INVESTIGATOR

I confirm that I have provided information and explained the nature and effect of the study to the subject and that his/her consent has been given freely and voluntarily

Signed: _____

Status: _____ Date: _____

Appendix 7.3 Patient consent

PATIENT CONSENT FORM

Study Title:

The Benefits of a Proficiency-Based Virtual Reality Simulation Curriculum in the Training of Future Providers of Colonoscopy in Ireland : Skills Transfer from the Virtual to the Real Environment.

Please circle the appropriate answer.

I confirm that I have read and understood the Patient Information Leaflet dated _____ attached, and that I have had ample opportunity to ask questions all of which have been satisfactorily answered.

Yes

No

*I understand that my participation in this study is entirely **voluntary** and that I may withdraw at any time, without giving reason, and without this decision affecting my future treatment or medical care.*

Yes

No

I understand that my records may be viewed by individuals with delegated authority from Beaumont Hospital or RCSI.

Yes

No

I understand that my identity will remain confidential at all times.

Yes

No

I have been given a copy of the Patient Information Leaflet

Yes **No**

FUTURE USE OF ANONYMOUS DATA:

I agree that I will not restrict the use to which the results of this study may be put. I give my approval that unidentifiable data concerning my person may be stored or

Appendix 7.3 Patient consent

PATIENT CONSENT FORM

Study Title:

The Benefits of a Proficiency-Based Virtual Reality Simulation Curriculum in the Training of Future Providers of Colonoscopy in Ireland : Skills Transfer from the Virtual to the Real Environment.

Please circle the appropriate answer.

I confirm that I have read and understood the Patient Information Leaflet dated _____ attached, and that I have had ample opportunity to ask questions all of which have been satisfactorily answered.

Yes

No

*I understand that my participation in this study is entirely **voluntary** and that I may withdraw at any time, without giving reason, and without this decision affecting my future treatment or medical care.*

Yes

No

I understand that my records may be viewed by individuals with delegated authority from Beaumont Hospital or RCSI.

Yes

No

I understand that my identity will remain confidential at all times.

Yes

No

I have been given a copy of the Patient Information Leaflet

Yes No

FUTURE USE OF ANONYMOUS DATA:

I agree that I will not restrict the use to which the results of this study may be put. I give my approval that unidentifiable data concerning my person may be stored or

electronically processed for the purpose of scientific research and may be used in related or other studies in the future. (This would be subject to approval by an independent body, which safeguards the welfare and rights of people in biomedical research studies - the Beaumont Hospital Ethics (Medical Research) Committee.)

Yes

No

Patient _____

Signature and dated

Name in block capitals

To be completed by the Principal Investigator or his nominee.

I the undersigned, have taken the time to fully explained to the above patient the nature and purpose of this study in a manner that he/she could understand. I have explained any risks/benefits and have invited him/here to ask questions on any aspect of the study that concerned them.

Signature:

Date:

Name in Block Capitals:

Qualification:

Appendix 7.4 – colonoscopy assessment

COLONOSCOPY PROCEDURAL ASSESSMENT FORM

Subject code: _____

Date of procedure: _____

Patient identifier: _____

Evaluator: _____

Difficulty of procedure: difficult average easy

Endoscopist:

	PREPARATION	yes	no	score
1	Did the subject check that the consent form was signed?			
2	Did the subject demonstrate awareness of the appropriate monitors?			
3	Did they demonstrate knowledge of the patient history? (ie look at medical chart)			
4	Did they check the equipment before commencing?			

Level: Consultant Trainee Year _____ (if trainee)

Handedness: left right

Approximate number colonoscopies performed: _____

	PROCEDURE	yes	no	score
1	Did they perform a PR exam before inserting the colonoscope?			
2	Did they respond appropriately to patient discomfort? (if occurred)			
3	Could they correctly identify pathology? (if present) NA <input type="checkbox"/>			
4	Could they correctly identify normal tissue?			
5	Procedure completion:			
	Did they reach the sigmoid colon?			(3)
	Did they reach the transverse colon?			(3)
	Did they reach the descending colon? (ie pass the hepatic flexure)			(3)
	Did they reach the cecum? (in 30 mins or less)			(3)
6	Was the procedure free from serious errors? (bleeding/perforation)			1
7	Did the supervising endoscopist have to take over (where the case was <i>not</i> inappropriately difficult)			(-1..)
	Number of times this occurred <input type="checkbox"/>			
8	Was it necessary to give an agent to reverse sedation?			1

Score for this section: _____ (max 23)

Section 2 - global

Global Assessment

variable	0	1	2	3	4
Respect for tissue	Often used unnecessary force, or caused damage by inappropriate use of colonoscope		Careful to avoid damaging bowel wall but occasionally caused inadvertent damage		Consistently handled tissue appropriately with minimal damage

variable	0	1	2	3	4
Aware of position, anatomy	Lack of awareness of position, frequently no view of lumen		Reasonably familiar with landmarks, occasionally requires direction		Recognises all landmarks, aware of location at all times. Good luminal view

variable	0	1	2	3	4
Ability to use colonoscopy	Many unnecessary or inefficient moves, difficulty controlling colonoscope, no torque steering, awkward		Generally familiar with all controls, few unnecessary moves, occasionally awkward		Demonstrates consist ease of use of instrument, appropriate use of all controls, fluid movements

Variable	0	1	2	3	4
Quality of examination	Frequently inadequate view of mucosal surface, high chance of missing pathology		Generally adequate visualisation of mucosa, occasional verbal instruction from supervisor		Adequate view maintained at all times, no further pathology encountered on withdrawal

Variable	0	1	2	3	4
Response to difficulties	Unaware when encountering difficulty, progresses scope regardless		Recognises difficulties, ie no luminal view, patient discomfort, attempts to resolve		Recognises and takes appropriate action when encounters difficulty

Variable	0	1	2	3	4
Knowledge and flow of procedure	Frequently stopped case, unsure of next move		Reasonably knowledgeable, occasional hesitation		Familiar with all aspects of procedure, easy flow

variable	0	1	2	3	4
Overall performance	Very poor		average		excellent

Variable	0	1	2	3	4
assessment of patient comfort	Generally poor, frequently in pain		Reasonably good, occasional discomfort		Pain-free and comfortable throughout

Appendix (i) Individual Thesis Chapter Abstracts

Chapter 1 - Background

This chapter provides a detailed background to the thesis. In addition to a brief history of surgery, there is a review of the three main procedures used in the studies – colonoscopy, hand-assisted laparoscopic colectomy (HALC) and endovascular renal artery stenting. Surgical training from historical times to the present is reviewed in order to contextualise the thesis, and there is a section relating to the area of simulation in general, and in surgical training. Finally there is a discussion of learning theories and their applicability to surgical training.

Chapter 2 - Materials and Methods overview

This chapter gives a description of the simulators and the aptitude tests which were used in all the studies.

Chapter 3 - Improving surgical training: feedback reduces error scores in simulated laparoscopic colectomy

Introduction: It is known that feedback enhances the learning process, although the optimal type and frequency of feedback is not established. We used the ProMIS Hand-assisted Laparoscopic Colectomy (HALC) simulator to train residents, and hypothesised that giving subjects feedback during a training session would improve their performance and learning curves.

Methods: We tested 16 residents (Group 1) who performed 5 HALC procedures on the ProMIS simulator. Efficiency of instrument use and pre defined intra operative error scores were assessed. Facilitators assisted their performance and answered questions when asked. A similar cohort of 12 trainee surgeons (Group 2) then performed the same 5 cases, but with standardised feedback and the chance to review errors after every procedure. Skills transfer was also assessed in Group 2. Data were analysed using SPSS version 15. Means were compared using Mann-Whitney-U tests, with a p-value of <0.05 taken as significant.

Results: Group 1 achieved better results for instrument path length (23,874 vs 39,086, $p = 0.001$ overall, not significant (ns) on trials 1, 4) and instrument smoothness (2015 vs 2567, $p = 0.045$, ns on trials 1,2 and 4)). However, Group 2 (feedback) performed significantly better with regard to error scores (14 vs 4.42, $p = 0.000$). In addition they demonstrated a smoother learning curve and their performance had reached a plateau by the fourth trial. The worse instrument scores are likely a function of a more thorough dissection and error free procedure. Partial skills transfer was demonstrated to a similar procedure.

Conclusion: In conclusion the provision of standardised proximate feedback was associated with significantly less errors and an improved learning curve. This may help to make training more efficient and improve patient safety.

Chapter 4 -The effectiveness of endovascular simulator training for novices: can learning take place in the absence of expert feedback?.

Objectives: Complex endovascular skills are difficult to obtain in the clinical environment. Virtual reality (VR) simulator training is a valuable addition to current training curricula but is there a benefit in the absence of expert trainers?

Methods: 18 endovascular novices performed a renal artery angioplasty/stenting (RAS) on the VIST simulator. All trainees had identical pre-procedural teaching..

They were divided into three groups:

- Group A (n=6, control) - no performance feedback
- Group B (n=6, non-expert feedback) - feedback after every procedure from a non expert facilitator
- Group C (n=6, expert feedback) - feedback after every procedure from a consultant vascular surgeon.

Each trainee completed RAS 6 times. Performance metrics included procedural and fluoroscopy time, contrast volume, accuracy of balloon placement, and handling

errors. In addition, all procedures were recorded and objectively assessed. Data were analysed using SPSS version 15.

Results: A clear learning curve was observed across the 6 trials. There were no significant differences between the 3 groups for procedural or fluoroscopy time or contrast use, but group C made fewer errors ($p = 0.004$). Post-hoc testing revealed significant differences between groups A and C ($p = 0.006$), and groups B and C ($p = 0.014$). Group C performed best for accuracy of instrument placement although the differences were not significant. For the objectively assessed errors, there were significant differences between Group A and B ($p = 0.002$) and Group A and C ($p = 0.000$) but not between Group B and C.

Conclusion: VR simulator training for novices can significantly improve general performance in the absence of expert trainers, but procedure-specific qualitative metrics are improved with expert feedback. Objectively assessed error scores were not significantly affected by expert vs non-expert feedback.

Chapter 5 – Can Wii improve surgical performance? – the effect of non-surgical skills training on laparoscopic surgical skills

Background: It has been suggested that abilities in non-surgical tasks may translate to the surgical setting, with video-gaming attracting particular attention due to the obvious similarities in the manual skills required, such as hand-eye coordination and screen mediated task execution. The aim of this study was to prospectively assign laparoscopic novices to receive a period of structured practice on the Nintendo Wii and compare their performance of basic laparoscopic tasks before and after this sessions with a group of control subjects.

Methods: 22 medical students with no prior laparoscopic or video game experience were recruited to the study. They were randomised into 2 groups

Group 1 – control

Group 2 – Wii

All subjects performed 2 physical (Bead Transfer and Glove Cutting) and 1 virtual laparoscopic simulated tasks on the ProMIS surgical simulator. Performance metrics were measured. The same tasks were repeated an average of 7 days later, and in between the 2 sessions, the subjects in the Wii group had structured practice sessions on the Wii video game

Results: Taken together, all subjects significantly improved their performance from session 1 to session 2.

For the Bead Transfer task, subjects in the Wii group performed better on session 2, but not significantly. The Wii group significantly reduced the number of beads dropped compared to the control group between session 1 and 2 ($p = 0.045$), but there was no other difference between the groups' improvement for the other performance measures (T, IPL, IS). For the fine dissection task, a similar pattern was observed. The Wii group performed better on all metrics in session 2, but the differences were not statistically significant, and there was no difference between the groups in the improvement from session 1 to session 2.

For the virtual task, there was no significant improvement between session 1 and 2.

Conclusion: The novice subjects demonstrated a steep learning curve between their first and second attempts at the laparoscopic tasks. Practising on the Wii was associated with a better performance on session 2, although the difference was not significant. This suggests that a more intensive practice schedule may be associated with a better performance and that training on non-surgical tasks may be a cheap, convenient and effective addition to current training curricula.

Chapter 6 - Retention of surgical skills

Background – It has been suggested that surgical skills which are learnt may decay over time in the absence of continuous reinforcement. This is particularly relevant in the setting of intensive skills courses where trainees may not have the opportunity to immediately practice the skills learnt. The aim of this study was to investigate skill retention in the absence of reinforcement.

Methods – 5 subjects who underwent endovascular RAS training were retested after a mean interval of 15.8 weeks. They were retested on one performance of the RAS. Simulator-based metrics were analysed. 8 subjects who underwent HALC training were retested after a mean interval of 14.5 weeks. They performed 2 repetitions of the HALC procedure at retention testing. Simulator metrics and objective assessment of errors were recorded.

Results – Good skills retention was observed for the RAS subjects, who demonstrated mean scores which were similar to the performance between trials 4 and 5 for the general metrics and trials 3 and 4 for the procedure specific metrics. Complete retention was observed for the error scores. For the HALC subjects, retention of skills was poor for instrument handling scores, but complete for error scores. This was the only metric for which proficiency was obtained.

Conclusion – Partial skills retention was observed after a period of 15 weeks for instrument handling scores, and appeared better for an endovascular procedure. Procedure-specific error performance did not decay. This may be due to the fact that proficiency was obtained for this metric.

Chapter 7 - The benefits of a proficiency-based simulator training curriculum for colonoscopy

Background – Colonoscopy is an important diagnostic and therapeutic procedure. The demand for colonoscopy is set to increase further and training resources are already limited. There is currently a lack of structured training and competency assessment.

Methods – We designed a complete proficiency based curriculum for colonoscopy training including web-based teaching materials, a brief didactic teaching session, and proficiency-based training on the GI Mentor simulator. 24 subjects were recruited and randomised to receive simulator-based training, or continue with current training. The simulator trained group required a mean of 21 repetitions to reach proficiency. All subjects were assessed on their performance of real colonoscopies using a standardised, validated assessment form combining a checklist and global assessment.

Results – 24 subjects were recruited, 17 subjects performed colonoscopies and 64 procedures in total were assessed. The simulator-trained group performed significantly better with regard to the specific checklist score ($p = 0.046$) and although they performed better on the global and overall score the differences did not reach statistical significance.

– proficiency-based training incorporating simulation is an effective, safe and efficient way to train novices, and leads to significant performance benefits. This curriculum template could be used to deliver training for many other procedures.

Chapter 8 Influence of innate abilities on surgical skills

Background – It has been suggested that innate, or fundamental abilities may affect both technical ability in surgery and the rate at which trainees improve. There may be a role for the measurement of these abilities in order to guide selection into surgical training programmes, but this area requires further investigation.

Methods – We used a standard battery of tests of visuospatial, perceptual and psychomotor ability. We measured these abilities in all the subjects who underwent simulator training and analysed the relationship between these abilities and the subjects' performance.

Results – Data for 73 subjects were analysed according to procedure performed. Overall there were few significant correlations. For the subjects who performed the endovascular procedure, psychomotor ability significantly correlated with some performance metrics. For the subjects who performed the HALC cases, the only ability to correlate with performance was perceptual ability.

Conclusions – Our data were limited by small numbers but overall it appears that the relationship between innate ability and surgical performance requires more definition. Particular domains of innate ability may be more relevant to particular procedures and this might prove useful for designing training curricula around individual trainee needs.

Chapter 9 Conclusions

In the current climate of changes and challenges to surgical training it is imperative that the benefits of training are maximised and much future training is likely to be based in the non-clinical environment of skills labs and simulators. I believe that the findings of this research are thus topical and relevant and that they may be used to continue to improve surgical training.

Appendix (ii)

Presentations and publications

Improving surgical training: optimising the learning curve for a simulated surgical procedure.

E Boyle, M Al-Akash, P Neary, A Gallagher, O Traynor.

Oral presentation at:

Sylvester O'Halloran annual surgical meeting, Limerick, March 2008

Towards continuous improvement in endoscopy standards-validation of a colonoscopy assessment form

Boyle E, Al-Akash M, Hill ADK, Patchett S, McNamara D.

Poster presentation at:

Irish Society of Gastroenterology Summer meeting, Killarney, June 2009.

The effectiveness of endovascular simulator training for novices: can learning take place in the absence of expert feedback?

Boyle E, O'Keefe D, Naughton P, Hill ADK, Mc Donnell C, Moneley D.

Oral presentation at:

Vascular Society Annual Meeting November 2009, Liverpool.

Improving surgical training; the use of feedback to reduce errors and improve retention during a simulated surgical procedure.

E Boyle, M Al-Akash, P Neary, A Gallagher, O Traynor

Accepted for oral presentation at

SAGES surgical conference, Washington DC, April 2010

Neary PC, Boyle E, Delaney CP, Senagore AJ, Keane FB, Gallagher AG. Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc.* 2008;10:2301-9.