THE EFFECTS OF PRACTICAL TRAINING METHODS OF DIFFERENT FORMS AND INTENSITIES ON THE ACQUISITION OF CLINICAL SKILLS

by

ELPINIKI LAIOU

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Unit of Public Health, Epidemiology and Biostatistics
College of Medical and Dental Sciences
The University of Birmingham
October 2009

ABSTRACT

Simulation holds enormous potential for medical education, where patient safety concerns have made practice on patients less acceptable. However, there is no unequivocal evidence of simulation training translating to improved performance *in vivo*. Therefore, the aim of this thesis is to add to the literature on simulation training by a) synthesising the current evidence on the effectiveness of simulation training in healthcare, b) investigating the effectiveness of different 'doses' of mannequin training in learning laryngeal mask airway placement and c) assessing the effectiveness of a simulation course on managing life threatening illness.

This thesis has added to the literature in the field of medical education a review of reviews of the evidence regarding the effectiveness of simulation training in medicine and surgery, and two RCTs evaluating different simulation training courses. The review of reviews highlighted that simulation training can be effective, but there was little consistent evidence across tasks or types of simulator. The two RCTs reported nil results, reinforcing that simulation alone is insufficient to ensure effectiveness. These results highlight the importance of recognising when simulation training is appropriate, how simulation interacts with other elements of a training programme and how the simulation can be made maximally effective.

DEDICATION

This thesis is dedicated to the loving memory of my brother John Laios (1983-2005).

ACKNOWLEDGMENTS

First of all I would like to thank Dr Celia A. Brown and Professor Richard J. Lilford, my supervisors, for their immense support and assistance. I feel very privileged to have experienced their working ethic and guidance and they will always be role models to me.

I would also like to thank Dr Thomas H. Clutton-Brock for his enormous support in carrying out the two randomised controlled trials in this thesis.

I am grateful to all the members of the 'Lilford Team' for their support and friendship. Among them, I would like to thank Dr Russell Bowater for his invaluable statistical help.

Special thanks to Dr Joanne Moysey, Dr Sue Sinclair, Dr James Geoghan, Dr Ritu Gupta, Dr Suneel Desai, Dr Julian Bion and Dr Griselda Cooper for their invaluable assistance in conducting the trial assessments.

I would like to thank my family for their endless love and care. In addition, a big thanks to all of my friends for keeping me sane through difficult times. I would also like to thank Vangelis for his support and for his help with the administrative side of my trials.

I would like to acknowledge support of this work through the MATCH programme (EPSRC Grant GR/S29874/01) although the views expressed in this work are entirely my own.

Finally, I would like to thank everyone who took part in my research and made this thesis possible.

Elpiniki Laiou

October 2009

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AMEE	Association for Medical Education in Europe
ALERT	Acute Life-threatening Events-Recognition and Treatment
ARICM	Anaesthesia, Respiratory and Intensive Care Medicine
CONSORT	CONsolidated Standards Of Reporting Trials
CPR	Cardiopulmonary Resuscitation
EOM	Economy of Movements
HDA	Health Development Agency
ILS	Immediate Life Support
IV	Intra Venus
LMA	Laryngeal Mask Airway
MBChB	Bachelor of Medicine
METI	Medical Education Technologies Inc
MOLTI	Management of Life-Threatening Illness
OR	Operating Room
OSCE	Objective Structured Clinical Examination
PBL	Problem-Based Learning
RCT	Randomised Controlled Trial
SP	Standardised Patient
VAM	Voice-Activation Mannequin
VR	Virtual Reality

PREFACE

A *simulation* is a device or exercise that enables the participant to reproduce or represent phenomena that are likely to occur in actual performance[1]. In Medicine, it can be used as an educational technique that allows interactive, and at times immersive, activity by recreating all or part of a clinical experience without exposing patients to the associated risks[2]. This holds enormous potential for medical education, where increasing concerns regarding patient safety have made 'learning by doing' less acceptable than in the past. However, there is no unequivocal evidence of simulation translating to *in vivo* effectiveness across the board of technical and non-technical skills in healthcare.

Therefore, the main aim of this thesis is to add to the literature on the effectiveness of simulation training through three specific objectives. Firstly, to identify and synthesise review-level material to highlight the current evidence on the efficacy and effectiveness of simulation in training health care professionals. Secondly, to investigate the effectiveness *in vivo* of different 'doses' of mannequin training in learning a simple practical procedure, insertion of a laryngeal mask airway (LMA). Thirdly, to assess *in vitro* the efficacy of an undergraduate simulation course on managing acutely ill patients using a computer-controlled mannequin. A secondary aspect of the latter two objectives is to compare self-assessment of success in the task at hand with objective assessment by a third party.

As a result this thesis is composed of the following sections. Chapter 1 is a general introduction of why and how simulation training could be used to enhance the learning of clinical skills, including broad applications of simulation in medicine and a brief discussion of different types of simulators. Chapter 2 presents the educational theories that can be used to

explain how simulation training leads to improved performance and highlights a number of methodological considerations with regard to the evaluation of simulation training. Chapter 3 documents a systematic review of reviews examining the evidence on the efficacy and effectiveness of simulation training of technical and non-technical skills on *in vitro* and *in vivo* performance of health professionals/trainees.

The subsequent two chapters are randomised controlled trials (RCTs) which focus on the evaluation of simulation training of different forms and intensities on the acquisition of clinical skills. Chapter 4 presents an RCT comparing the effectiveness *in vivo* of two LMA placement mannequin simulation courses of different durations. Chapter 5 reports an RCT assessing the efficacy of the management of life-threatening illness (MOLTI) course, a Bachelor of Medicine (MBChB) Year 5 simulation course at The University of Birmingham, using the SimMan® simulator. This study has been presented at the *AMEE 2008* conference. Chapter 6 is the last chapter of this thesis and presents an overall synthesis of the theory and research reviewed and undertaken as part of this thesis along with implications for simulation training practice and suggestions for future research.

CHAPTER 1

GENERAL INTRODUCTION

1.1. CHAPTER OVERVIEW

The present chapter aims to provide an outline of how and why simulation training could be used to enhance the learning of clinical skills, which previously has relied on the traditional apprenticeship model. Both types of training are discussed in terms of rationale, advantages, disadvantages and how they can fit together towards fostering the attainment of clinical skill. Simulation training is also discussed in terms of broad applications and types of simulators.

1.2. THE APPRENTICESHIP MODEL

For thousands of years traditional clinical practice learning has relied on the apprenticeship model[3]. The apprenticeship model is a form of learning by doing, where a trainee learns how to perform a task under close guidance from an expert[4]. In this model, surgical skills, for example, are taught by the student directly observing and then imitating the actions of an expert mentor with 'see one, do one, teach one' as the principal method of knowledge acquisition[5]. The surgical trainee gradually builds up the knowledge, clinical judgement and operative skills needed for independent practice over the course of many years[6].

1.2.1. ADVANTAGES OF THE APPRENTICESHIP MODEL

The invaluable advantage of the apprenticeship model is that it offers trainees *real life* exposure to both diseases and treatments[3]. It gives trainees the opportunity to experience the input of all those taking part in a normal working ward, and to practise physical examination, communication skills, interpersonal skills and practical procedures[7]. Taught in the context

within which each task is most functional (i.e. during patient encounters), the trainee comes to understand the complexities, difficulties and customs common to that environment, which are important to their ability to perform the task efficiently[4].

1.2.2. DISADVANTAGES OF THE APPRENTICESHIP MODEL

The apprenticeship model is unstructured and requires a high volume of cases, multiple opportunities for repetition, skilled mentors and long work hours[5, 8]. Some clinical procedures are so rare that they are difficult for trainees to 'see and do'[9]. Furthermore, the learning needs of the trainee are inevitably secondary to the medical care needs of the patient, making learning an opportunistic process[6]. The need for emergency care creates a poor context of learning in real life as the learner is often moved to an observer role in the midst of crisis[10]. Yet, the apprenticeship model is based on the assumption that in the course of their clinical practice trainees will encounter all types of disease and treatment and in sufficient volume to achieve competency[3].

In addition, the apprenticeship model assumes that everyone receives the same education, which is not the case due to variability in patient diagnoses, volumes, mentors' schedules and match of the mentors' skills to patient disease[3]. Moreover, repeated observation while assisting a mentor means that learners may engage in parts of the procedure that are too easy or too difficult to result into meaningful learning, while immediate feedback may not always be possible[11]. Consequently, the apprenticeship model does not deliver learner-centered training.

In modern medical practice these issues are further compounded by work hours restrictions that reduce the exposure of junior doctors to their mentors[5]. The European Working Time Directive requires a maximum 48 hour working week for doctors in training starting on 1st August 2009[12]. Furthermore, the increase in medical school numbers in the UK has ensuing consequences for clinical training at undergraduate and postgraduate level[13]. These include increased demands in terms of the number of mentors and patients required, a subsequent increase in patient safety concerns, and potential increase in the duration of clinics to accommodate student-patient interactions. Consequently, it becomes important for students and trainees to have good foundations of basic practical and interpersonal skills prior to entering the clinical environment in order to maximise clinical placement learning opportunities[2].

1.2.3. REASONS LEADING TOWARDS SIMULATION TRAINING

Due to the issues discussed above, it appears that the apprenticeship-style training in medicine can no longer keep up with changes in health policy, medical training and culture. In summary, the drivers to learning through clinical simulation include:

- The failure of traditional learning modes[14].
- The increasing numbers of students[2].
- Changing clinical experiences[14]. Opportunities for surgical training have decreased because patterns of disease have changed with some diseases becoming rarer[1], e.g. there are fewer acute infections[7]. Further reasons for this change are an increasing number of patients who are hospitalised and shorter lengths of hospital stays[15].
- Shorter time in postgraduate training[14].
- Working time restrictions[2, 14].

- Patient safety[2, 14, 16]. The concept of "learning by doing" is less acceptable today particularly in the case of invasive and high-risk procedures[9].
- Financial constraints around training students and junior doctors in the operating room[17].
- Clinical governance[14]. There is a change of focus from process to outcome evidence of training success by accreditation councils[17].
- Widening participation in learning and enhancing the learning environment for healthcare professionals engaging in training and continued professional development[18].
- Professional regulation agendas of team-based learning and interprofessional learning[14].
- New technologies for diagnosis and management[16].

Simulation training is increasingly viewed by medical educators and accreditation councils as holding great potential towards addressing the abovementioned problems. The UK General Medical Council recommends the use of new technologies, including simulation, by medical schools to deliver teaching, notes the importance of experiential learning in simulated clinical settings and advocates that students must be given opportunities to develop and improve their clinical and practical skills in skills laboratories before they use these skills in clinical situations[19]. In addition, advances in technology are making more realistic simulations possible.

1.3. SIMULATIONS IN MEDICINE

Simulations originated in nonmedical settings and have been used as a teaching tool in aviation, the military, the nuclear power industry and business for the past three decades[9, 20]. Simulations aim to duplicate the essential elements of reality in a controlled manner[20]. It is thought that the use of simulators in medicine could provide trainees with both initial background information on technique and indications for procedures and early hands-on training experience to shorten the initial *in vivo* critical learning curve and accelerate clinical practice learning[21]. However, in contrast to aviation, medicine has few constraints or givens - pilots fly only specific types of planes, use very tightly controlled procedures and deal with limited numbers of crises. The human body is not as predictable as a plane; different patients having the same disease may present different symptoms and may respond differently to the same treatment (see section 1.3.4). Consequently, in medicine the number of different situations to be simulated is much greater, which may make simulation training less feasible as a comprehensive form of learning.

1.3.1 APPLICATIONS OF SIMULATION IN MEDICINE

Simulation training has been extensively used in the medical fields of anaesthesiology, critical care, cardiology and open and laparoscopic surgery[10]. Examples of the applications of simulation training in different medical skills and across the various medical fields can be found in Chapter 3 of this thesis, which is a review of reviews on the efficacy and effectiveness of simulation training in Medicine.

Simulations can provide trainees with an opportunity to develop psychomotor and non-technical skills, allowing them to have their first encounters with real patients when they are at higher levels of technical and clinical proficiency[22]. Thus, the aim of simulation in this case is to improve the performance of the participating trainees, with an ultimate endpoint the decrease of patient mortality and morbidity. The various types of simulation training seek to improve one or more aspects of performance, including technical skills (e.g. catheterisation), cognitive skills (e.g. decision-making, planning, situation awareness), and social/interactive skills (e.g. communication, team-working, leadership)[23, 24]. The educational theories underlying simulation training will be presented in section 2.3.

Furthermore, simulations allow for formative assessment, which is an important element of deliberate practice[2]. The majority of assessments target basic psychomotor and communication skills. However, as the skills become more complex, the assessment challenge increases[2]. Other applications of simulation include summative assessment of competency, engaging in research and analysis, modelling disease processes and treatment outcomes, and aiding, training and rehabilitating patients[20]. In particular, assessing professional competence is one of the factors contributing to the rise of simulations in medical education, as simulation technology allows the provision of standardised experiences for all examinees, using learner-specific findings and reliable outcome measures[16]. The most common simulation format in assessment is the objective structured clinical examination (OSCE)[25].

With regards to simulations used for testing theoretical models, events and processes can be simulated on a computer. For example, a theory may be developed regarding a) what kind of

challenges the medical work-a-day world would provide for the practitioner and b) how that person would handle specific challenges[20]. In order to check the accuracy of that theory, a software program can be developed based on the theory's characteristics and used to compare the output of the program against real practitioners undergoing the same challenges[20].

Computer simulations of the impact of medical treatment on patients' physiological response can also be used to develop step-by-step theories about antecedent causes of medical conditions[20]. Furthermore, simulators can be used to assess medical equipment, information systems and procedures[9, 22]. Simulations can be used to explore vulnerabilities in health care delivery and improve the competence of providers and the system of care[22]. Simulators can also be used in surgical planning; planned procedures on a specific patient can be first rehearsed on a simulator by installing in it their anatomical details on radiological imaging[8]. Another example is programs that aid surgeons during operations by allowing them to see the internal anatomic structures based on the patient's radiographic study that are superimposed on the surface anatomy[26].

With regards to aiding patients, virtual reality can be used to desensitise patients with severe phobias or post traumatic stress disorder through simulation of high-stress situations[9]. It becomes apparent that simulation can be applied to a broad range of applications in medicine, with training simulators and their effectiveness constituting the focus of the research presented in the following chapters. A categorisation of the types of training simulators within the scope of this thesis is presented below.

1.3.2. TYPES OF TRAINING SIMULATORS

Several simulator categorisations can be found in the medical training literature[1, 2, 6, 22, 27-30]. Alinier proposed a typology of simulation methodologies in six technological levels with level zero being the lowest[27] (Table 1.1).

Table 1.1. The Alinier typology

Simulation technique	Mode of delivery	Type of Skills addressed	Typical use
Level 0 Written simulations (pen and paper simulations or 'Patient Management Problems' and latent images)	Usually student led	Passive cognitive	Patient management problems Diagnosis Mainly for assessment
Level 1 3-D models (basic mannequins, low-fidelity simulation models, part-task simulators)	Student or trainer led	Psychomotor	Demonstration and practice of skills
Level 2 Screen-based simulators (Computer simulation, Simulation software, videos, DVDs, or Virtual Reality and surgical simulators)	Student or trainer led	Interactive cognitive	Cognitive skills Clinical management Sometimes interpersonal skills (if team interacting over net-worked computers)
Level 3 Standardised patients (Real or simulated patients (trained actors), Role Play)	Student or trainer led	Psychomotor, cognitive, and interpersonal	Cognitive skills Clinical management Interpersonal skills Physical assessment Diagnostic or management problems
Level 4 Intermediate fidelity patient simulators (Computer controlled, programmable full-body size patient simulators not fully interactive)	Trainer led	Psychomotor, cognitive, and interpersonal	Cognitive skills Clinical management Interpersonal skills Physical assessment Diagnostic or management problems Procedural skills Full-scale simulation training Demonstrations
Level 5 Interactive patient simulators (Computer controlled model driven patient simulators also known as high fidelity simulation platforms)	Preferably student led	Psychomotor, cognitive, and interpersonal	Cognitive skills Clinical management Interpersonal skills Physical assessment Diagnostic or management problems Procedural skills Full-scale simulation training Demonstrations

The Alinier typology has been provided in detail in this section as it is one of the most comprehensive and useful categorisations encountered in the literature and could serve medical educators in determining the appropriate type of simulation tool for the skill to be taught. Other simulator categorisations in the literature are:

- Beaubien & Baker proposed a simple typology of three categories including case studies/role play, part-task trainers and full mission simulation[28].
- Cumin & Merry proposed a classification system for simulators used in anaesthesia based on three attributes; interaction (hardware-based, screen-based or virtual reality-based), physiology (no physiology, script-controlled or model-controlled) and use for teaching (knowledge, cognitive skills or psychomotor skills)[29].
- Kneebone divided simulators into model-based (i.e. those based on physical models), computer-based (i.e. those that use computers to create illusions of reality, including virtual reality) and hybrid (i.e. those combining physical models with computers)[6].
- Meller proposed a typology of simulators for medical education including four elements represented as four 'P's (the patient and/or their disease process, the procedure or diagnostic test or equipment being used, the physician or paraprofessional and the professor or expert practitioner) with each element being further classed as passive, active or interactive[30].
- Torkington et al provided a categorisation for surgical simulations that included i) inanimate artificial tissues and organs ii) fresh tissue or animal models iii) virtual real and computerised simulation and iv) actors role-playing a trauma simulation[1].
- Ziv et al divided simulation tools and approaches to five main categories including low-tech simulators, simulated/standardised patients, screen-based computer

simulators, complex task trainers (including virtual reality) and realistic patient simulators[22].

The categorisation by Maran and Glavin[2] was selected for this thesis for being the most inclusive for the scope of this thesis, descriptive with regards to different types of simulators, simple and with minimum overlap between the different types. Three more types of simulation have been added for the purposes of this thesis (human cadavers, animal models and bio-simulation models) as they are used for surgical and anaesthesia training. This categorisation is used to structure the evidence considered in the review of reviews in Chapter 3. Another aspect to be taken into consideration when categorising simulators is their fidelity - that is 'the extent to which the appearance and behaviour of the simulator/simulation match the appearance and behaviour of the simulated system'[2]. However, the fidelity of a simulation is dependent on the manner in which the simulator is being used and on the type of task. For example, an integrated simulator has multiple features and applications that may or may not be used during a clinical management scenario. Using the categorisation of Maran and Glavin the different types of simulators can be broadly categorised to:

Part-task trainers: These low-fidelity simulators replicate only part of the environment and often resemble discrete anatomical areas of the body and are used to train basic psychomotor skills[2]. Examples include head and neck models for teaching airway management or bronchoscopy, such as the mannequins used in the RCT in Chapter 4, torso models for teaching central line placement, chest tube insertion or cardiopulmonary resuscitation, and forearm models for teaching peripheral intravenous or arterial line placement[31]. In addition, there are more sophisticated part-task trainers such as Harvery, a high-fidelity cardiovascular

system designed to help learners recognise common auscultatory cardiac findings[14]. Video-box trainers, a type of minimally invasive surgery simulator, can also be categorised as part-task trainers. These are opaque boxes that approximate the size of an adult human abdominal cavity that use real surgical instruments, video monitors, cameras and laparoscopes[32]. Their fidelity could be classed as low to medium.

Computer-based systems: Computer-based systems model aspects of human physiology or pharmacology, simulated tasks or environments and can provide feedback during or after the interaction[2]. Their interface is usually screen-based, lacking a body part mannequin component. Their fidelity could be classed as low to medium.

Virtual reality systems: Often used in combination with part-task trainers to present three-dimensional objects or environments to all human senses and are used extensively for endoscopic and laparoscopic training[2, 31]. Virtual reality has been described as 'a concept of advanced human-computer interaction that can be separated into five categories: immersive, desktop, pseudo, inverse and augmented reality'[9]. Their fidelity could be classed as medium to high.

Simulated patients: Used for teaching communication and interpersonal skills and are also used in combination with part-task trainers to increase the psychological fidelity of psychomotor skills training[2]. The term 'simulated patients' is interspersed with the term 'standardised patients', but the latter refers to actors, lay persons, healthcare staff or actual patients 'who have been coached to present their illness in a standardised way'[33]. Their fidelity is usually classed as medium to high.

Integrated simulators: Combine part or whole body mannequins with computers, which produce physical signs and feed physiological signals to monitors. These simulators can be driven by instructors who adjust signs to reflect patient responses or by scientifically derived complex mathematical models of respiratory and cardiovascular physiology, and extensive pharmacological modelling[2]. The term *patient simulators* is also used, which refers to 'full-scale, computer-controlled mannequins with programmable physiologic responses that can be used to present different lifelike disease states'[31]. Patient simulators such as the SimMan, which was used in the RCT in Chapter 5, and the Human Patient Simulator are integrated simulators. Their fidelity is usually classed as medium to high.

Simulated (working) environments: Can be used for team training and increase the psychological fidelity of scenarios by including high fidelity integrated simulators[2]. These can include whole operation rooms or patient wards, with a full range of the health care personnel participating as actors or trainees. Their fidelity is usually classed as high.

Furthermore, **human cadavers**, **animal models** or **animal tissue** may be used for surgical and anaesthesia training. Their fidelity could be classed as medium to high. However, cadavers are strictly anatomic simulators of the human body, lacking physical signs, are expensive and not always in supply[10]. The Cruelty of Animals Act of 1876 forbids the use of animals in surgical skills training in the UK[1] and in other countries upkeep cost issues are also prohibitive[32]. However, the use of animal parts is allowed in teaching surgical skills such as suturing and anastomosis and in the case of **ex-vivo/bio-simulation models** animal parts are placed in laparoscopic training boxes for practice of laparoscopic procedures[1].

1.3.3. ADVANTAGES OF SIMULATION TRAINING

Research in other domains has suggested that the attainment of true expertise requires 10,000 hours of deliberate practice (e.g. violin) and perhaps 10 years of commitment to the field[34]. Streufert et al stated that '*How* medical personnel handle problems whenever there is no easy 'right' answer not only involves content knowledge – it involves process of thought'[20]. Simulations can focus on content or process or, in some cases, on both[20], which is not always possible during real-life clinical practice. Medical educators, including clinicians, identify numerous advantages in using simulation training.

The main advantages of simulation cited in the literature are:

- Training poses no risk of harm to patients[35].
- Learners can engage in repeated practice of scenarios and actions[35].
- Learners are allowed to make mistakes, explore the consequences of their actions, react to rectify deviations and learn from these[22, 35].
- Learners can practice rare or infrequent events so training can offer exposure to 'patients' with all the disease states that health professionals need to be able to recognise and manage[3, 22].
- Team training and crisis management can be practiced in a controlled environment using a full range of the health care personnel involved in such cases[35].
- Training times can be real or altered to suit the training needs of the learners[35].
- The level of complexity can be altered to suit the training level of the learner[35].

Hence, simulation training allows for learner centred training that bypasses the disadvantages of the apprenticeship model discussed in section 1.2.2. It can lend support to characteristics that lead to effective learning such as allowing for uninterrupted feedback, repetitive practice, a range of training levels, clinical variation, a controlled environment and individualised learning, which are presented in more detail in section 2.4. As a result, trainees may be able to achieve competency prior to engaging in real-life patient care[3].

1.3.4. DISADVANTAGES OF SIMULATION TRAINING

The following limitations have been proposed with regards to simulation training:

- Human beings differ in response to even relatively simple interventions, and therefore, the creation of simulated 'pathophysiology' is subject to the biases and interpretation of the persons writing the scenarios in terms of the expected pathophysiology and response[2].
- Subtle clinical clues used in clinical practice such as changes in facial expression, muscle tone and skin are not replicated with present-day technology[2].
- Where the simulator cannot properly replicate the real-life tasks or task environment targeted by simulation training, clinicians might acquire inappropriate behaviours referred to as 'negative training' or develop a potentially dangerous false sense of security in one's skills[27, 36].
- Participants may have trouble suspending disbelief. Thus, cavalier behaviour may occur as the learner knows that no human life is at risk or hypervigilance may be present, as the learner knows that an event is about to occur[37].
- There are concerns that using technology in the practice and teaching of medicine may have dehumanising effects, detracting from the caring side of medicine[38].

- Trainees may not appreciate the emotional effects of acute real-life encounters that can affect their thinking abilities and skills in real circumstances and may become skilful at dealing with the training technology rather than with actual patients[27].

Thus, it has been argued that trainees should use simulation-training as an adjunct to a wide variety of different delivery methods towards teaching a particular skill and not as the sole method[27]. Furthermore, medical educators should always bear in mind that the simulator is not an instructor, nor a curriculum, but rather a tool that can be used by a teacher towards content delivery[3]. It has been noted that 'simulations are often accepted uncritically, with undue emphasis being placed on technological sophistication at the expense of theory-based design'[11]. The theoretical framework underlying simulation training and its ability to facilitate learning is further discussed in Chapter 2.

1.4. SUMMARY

This chapter has highlighted the weaknesses of the apprenticeship model and how these lead towards the adoption of simulation training as an invaluable adjunct to medical training. The applications of simulation in medicine have been presented along with the different types of simulators. A range of different categorisations of simulators available in the literature have been outlined and the categorisation applied throughout this thesis has been presented. Based on the latter, simulators are broadly categorised to part-task trainers, computer-based systems, virtual reality systems, simulated patients, integrated simulators, simulated environments, human cadavers, animal models and bio simulation models. Simulation fidelity has been defined as the extent to which the appearance and behaviour of the simulation matches that of the simulated system with part-task trainers usually found on the lower end of the fidelity

spectrum and simulated environments on the higher end. The advantages and disadvantages of simulation training have been presented. Overall, simulators appear to hold enormous potential for medical training as a tool towards content delivery, but educators should bear in mind that these tools should be used only when appropriate and in accordance with the theories underlying learning.

CHAPTER 2

THE THEORETICAL BACKGROUND OF MEDICAL SIMULATION TRAINING AND ITS EVALUATION

2.1. CHAPTER OVERVIEW

The present chapter aims to describe the causal chain from simulation training to health outcomes and provide an outline of the educational theories that can be used to explain how simulation training leads to improved performance and of the characteristics of simulation training that enhance clinical skills learning. Higher education course design principles are briefly presented as a guide to simulation training design. The evaluation of simulation training is then discussed in terms of methodological considerations.

2.2. A CAUSAL CHAIN FROM SIMULATION TRAINING TO HEALTH OUTCOMES

Skills learned on a simulator should transfer positively both between differing levels of simulation and from the simulator to the real life[2] as shown in Fig. 2.1. That is, effective simulation training (A) should lead to improved performance on the simulator (B) and improved real life clinical performance (C). Improved clinical performance should then lead to improved patient outcomes (D). A definition of skills is 'actions (and reactions) which an individual performs in a competent way in order to achieve a goal'[38]. Transfer of training has been described as 'the extent to which knowledge, skills and attitudes acquired in a training program are applied, generalized, and maintained over some time in the job environment'[24].

Performance measures can refer to specific assessments of individual components of the performance or global measures rating the entire performance.

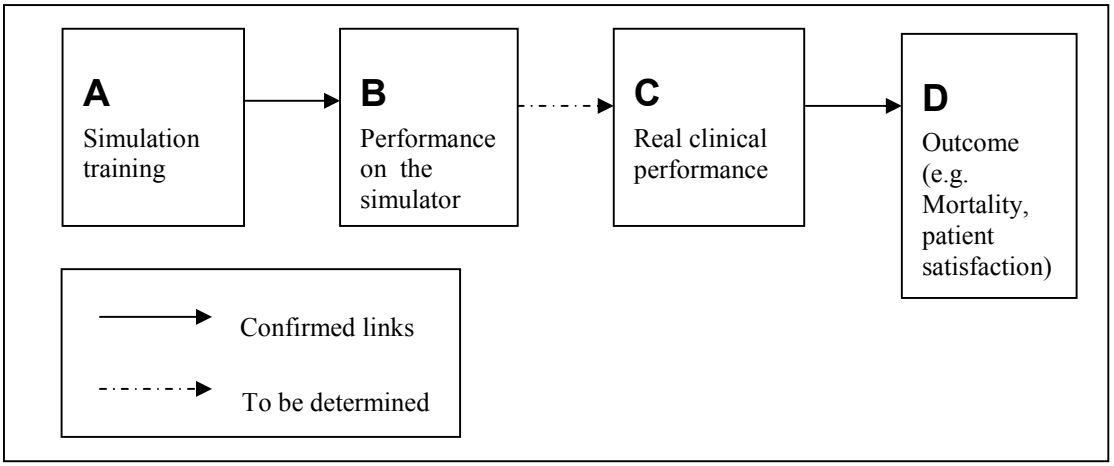

Figure 2.1. A causal chain from training to outcome in the evaluation of simulation

While measuring patient outcomes (D) is the ideal outcome to determine the effectiveness of simulation training (A), due to practical and scientific considerations measurement of clinical performance in real life (clinical process) (C) can be used as a surrogate outcome[36, 39]. This is acceptable if the link between the relevant processes and outcome has been established: if C leads to D and B leads to C then B can also lead to D (Fig. 2.1). Under the same logic, if B leads to C and C to D, then finding an improvement in B following simulation training should be sufficient evidence of effectiveness of A. A number of simulation training interventions have been evaluated in non-clinical settings with the results confirming the link from A to B[40, 41]. In addition, many clinical processes have been shown to be effective in improving health outcomes[1], although this is not universal (C to D). The link showing that improved training outcomes *in vitro* translate to improved clinical processes *in vivo* (B to C) remains to be determined for simulation[36, 42, 43]. Providing evidence to support (or refute) this link is one of the key themes of this thesis. Current evidence for this link will be reviewed in Chapter 3

while the following section is an outline of the educational theories that have been associated with simulation training and its potential to lead to learning and improved performance.

2.3. HOW SIMULATION TRAINING WORKS IN THEORY

Educational theories aim to explain the process(es) by which an educational experience translates into learning and improved knowledge and/or performance. Simulation training is compatible with several educational theories, which will be summarised in the present section (i.e. the characteristics of simulation training suggest that trainees may learn in the manner suggested by the theory). These theories also underpin the apprenticeship model, the traditional clinical practice 'learning by doing' model, which has been presented in section 1.2. Understanding the theories underlying simulation training is important in order to make sense of how simulation has the potential to help various learning processes to take place.

The first educational theory is Kolb's theory of experiential learning[4, 37]. According to the *experiential learning theory*, learning is defined as 'the process whereby knowledge is created through transformation of experience' with knowledge resulting from a four-stage recurring learning cycle where 'concrete experiences' become the basis for 'reflective observation' leading to 'abstract conceptualisation' and new concepts subsequently undergoing 'active experimentation'[44]. Simulation training can provide learners with these concrete experiences and opportunities for active experimentation through practice on the simulator, however, the value of the concrete experiences and opportunities provided by a simulator may depend on the degree of fidelity compared to the simulated activity.

Reflective practice emphasises supporting the learner's process of reflection during planning actions, acting, evaluating and re-conceptualising following an experience[45]. The learner's theories, contextual information and values can all influence each of these stages of reflection[45]. Again, practice on a simulator can provide the experiences that learners need to reflect on in order to achieve learning but the degree of support of the learner's reflection process may depend on specific features of simulation such as the availability of feedback, the time available for reflection and the fidelity of the simulation.

Constructivism describes how learners use their interactions with the world to construct their understanding of it, by assimilating experiences that fit into their existing cognitive structures and modifying their cognitive structures to accommodate experiences that do not fit[45, 46]. This learning environment needs to be perceived as safe by the learner so that preconceptions that could lead to disengagement can be exposed without fearing ridicule or injury to the patient[45]. Simulation training can provide such a safe and interactive learning environment, though this benefit could be seen as dependant on the actual effectiveness of simulation training in resulting in a valid understanding of the real life clinical process.

Social constructivism emphasises the importance of the social interaction in helping learners to construct new understanding[45]. During group simulation exercises the learner can advance within a 'zone of proximal development' with the instructor providing a 'scaffold' to learning in the initial stages and then gradually withdrawing that support to encourage learner independence[11, 45, 46]. This theory applies only partially to simulation training as an instructor is not always an inherent part of medical simulation training.

Simulation training is also compatible with the *situated learning theory*, which is also very akin to the apprenticeship model. According to this theory, learning takes place through 'legitimate peripheral participation' in a community of practitioners. That is, learners initially become engaged through secondary participation in the practices to be learned and gradually move towards more and more central forms of participation[11, 45, 46]. For example, a trainee surgeon initially only assists their mentor in conducting an operation and gradually moves towards undertaking the whole procedure. Simulator experience can be seen as legitimate peripheral participation where authentic learning experiences can be offered to students unconstrained by issues of patient safety[45], however, this requires the presence of an instructor which is not always the case with medical simulation training.

Furthermore, simulation training is compatible with the *humanist theories* that describe learners as being responsible for their own learning and the process of learning being about developing the individual rather than individual competencies[6, 46]. Simulation exercises allow learners to have control over clinical scenarios, which promotes self-evaluation that can be re-enforced by feedback[46]. Simulation training can also fit with *behaviourist theories.* The behaviourist theories focus on the stimulus- response principle and emphasise the importance of consistent positive and negative feedback in developing a skill, while they ignore the impact of the higher level internal processes of the mind in the transfer of learning to wider applications[6, 45, 46], which is the case in most real life clinical processes. Learners engaging in simulation training can receive such consistent feedback by their instructor and in the case of high-fidelity simulations, by the simulator itself.

The *Activity theory* emphasises the learning that occurs within different activity systems in which a learner operates and across the boundaries that exist between these systems[45]. Each activity system consists of individual and group actions directed towards an object[45]. For example, simulation could be seen as an activity system whose object is the learning of the trainee. The corresponding 'real world' practice could be seen as another activity system whose object may be sliding between trainee learning and the patient's needs for ethical medical management[45]. Related activity systems can be used jointly to construct meaning about objects on their boundary and to come to jointly shared understanding about these, resulting in enhanced learning[45]. This theory could be used to explore the problems arising when developing a skill in the simulator and using it in the operating theatre[45],however, this theory could also be seen as having limited practical value in creating concrete hypotheses regarding simulation training effectiveness.

While simulation training appears to fit to some extent with a number of educational theories of learning, considering the validity of these theories is beyond the scope of this thesis. These theories are not mutually exclusive and each may be useful in understanding different components of medical learning. For example, *behaviourist theories* could be seen as relating to the acquisition of specific manipulative skills while *constructivism* may be helpful in understanding the development of clinical judgement[6].

2.4. CHARACTERISTICS OF SIMULATION THAT FACILITATE LEARNING

As mentioned in the previous chapter, a simulator is only part of a training programme, a tool that can be used towards content delivery. The simulation training itself is made up of a number of characteristics. The effectiveness of simulation training in facilitating learning will depend on the presence and nature of these characteristics and, in some instances, on the interactions between characteristics. The desirable simulation characteristics for effective learning are described below and many of these also constitute the advantages often cited for simulation, mentioned in Chapter 1. These characteristics can be linked to the educational theories described above. However, it is important to recognise that many of these characteristics are tenets of good educational practice *per se* and are not limited to simulation training. According to a systematic review by Issenberg et al[16], which will be described in section 3.3.3, the weight of the best available evidence from simulation training research suggests that medical simulations are most effective in facilitating learning when the following characteristics are present:

Feedback

As mentioned in section 2.3, behaviourist theories emphasise the importance of feedback in developing a skill and in supporting reflective practice. Knowledge of one's performance through focused constructive feedback, which can be offered as part of simulation training, is considered a key component in clinical skill acquisition and maintenance[16]. Feedback allows learners to self-assess and monitor their learning progress and can slow the decay of acquired skills[47]. It encourages learners to reflect on their performance in ways that are missing from everyday clinical practice alone[42]. The presence of feedback is more important than its source, which can be built into the simulator or stem from the clinical teacher[47].

Repetitive practice

Repetitive practice can be linked to behaviourist theories, which focus on the stimulus-response principle[6]. Repetitive practice can also be linked to experiential learning theory since Kolb's four-stage cycle of learning described in section 2.3 is a recurring one. Skill repetition should be an essential feature of simulations giving learners the opportunity to engage in focused practice, where the aim is skill improvement[16]. Through engagement in intense repetitive practice, learners detect and correct errors and hone their skills making their performance automatic[47]. A term often used to describe the type of repetitive practice that learners should aim for is *deliberate* practice. Deliberate practice involves engaging in repetitive performance of cognitive or psychomotor skills in a controlled setting, which is of an appropriate difficulty level for the particular individual and includes rigorous assessment and provision of valid informative feedback[16, 16, 38, 48]. Deliberate practice should be focused on a well-defined area allowing the learner to correct errors and improve performance[46]. Simulation can allow for repetitive practice of technical, cognitive and interactive skills individually or simultaneously[46]. As mentioned in section 1.3.3, the acquisition of expertise requires many years of sustained *deliberate practice*[11, 48].

Curriculum integration

Curriculum integration appears to be supported by the activity theory presented in section 2.3. Simulation and the curriculum activities could be seen as related activity systems whose common object is the learning of the trainee. Based on the activity theory, the integration of simulation into the curriculum will result in jointly shared understanding about the learning objects on their common boundary and thus, in enhanced learning. In order to be most effective, simulation training should be built into learners' normal training schedule in order

to lead to learner engagement in deliberate practice, as optional exercises can be seen with less interest by the learner[16, 38, 47].

Range of training levels

Offering a range of difficulty levels starting at basic level and proceeding to training at progressively higher difficulty levels is seen as an important variable in simulation training[16]. Learning is enhanced when trainees have the opportunity to practice a task across a wide range of difficulty levels demonstrating performance mastery in each level against objective criteria and standards[47]. This characteristic could also be seen as fitting with the experiential learning theory which views learning as a recurring cycle where new concepts, in this case progressively higher difficulty levels, are constantly put into active experimentation (see section 2.3).

Multiple learning strategies

Simulation training should be adaptable to multiple learning strategies such as instructor-centred large group teaching, instructor-centred small group teaching, small-group independent leaning without an instructor and individual independent learning[16]. The choice of strategy is determined by the desired outcomes, available resources and educational culture of each institution[47]. The educational strategy chosen should match the educational goals and the extent of prior learning among trainees[16]. The use of multiple learning strategies means that most of the educational theories described in section 2.3 can come into play giving trainees the opportunity to experience the modes of learning that are ideal to them as individual learners.

Individualised learning

Simulation can offer learners individualised educational experiences adapted to each one's unique learning needs, where they are active participants[16]. Thus, learners can have complex tasks broken down into their component steps and progress in sequence and at their own pace towards defined educational outcomes[16]. This in turn enables controlled, proactive clinical exposure of learners to gradually more complex clinical scenarios including clinical variation[49]. Individualised learning is a central concept of humanist theories (see section 2.3).

Clinical variation

The more useful simulations can represent a wide variety of patient problems, patient demographics, and responses to treatment and thus increase the variety of patients that learners encounter, including rare, life-threatening patient problems[16]. This feature is critical for developing problem-solving skills but is not as critical for simulators designed for a specific task such as carotid stent placement, which have a narrower focus[47]. Providing clinical variation through simulation also helps standardise a clinical curriculum across different sites as it ensures that everyone is getting an equal variety of cases compared to trainees simply learning from the cases that happen to be present in their clinical environment[47]. Clinical variation promotes learning through the process of re-conceptualisation, which has been described in the experiential learning theory, reflective practice and constructivism (see section 2.3).

Controlled non-patient-facing environment

Simulation training can offer learners the chance to practice in a controlled environment where they can make errors without adverse consequences and instructors can focus on the learners rather than the patients[16]. Thus, in simulation training the allowance of errors becomes a very powerful educational tool in saving human lives[49]. This characteristic offers learners the opportunity to practice in an environment that could be perceived as safe, which promotes engagement in learning, as recommended in constructivism (see section 2.3).

Defined outcomes

Simulation training outcomes should be clearly defined as learners are then more likely to master the targeted key skills[16, 47]. This characteristic is important for self-evaluation and could be seen as an important prerequisite towards achieving learning through reflective practice, for the reflective observation stage of Kolb's cycle and for the humanist theories approach (see section 2.3).

Fidelity

Medical simulations should generate the autonomic, cognitive, and behavioural responses seen in the real-world and fidelity is important in order to create participants' experience of absolute realism[20]. A definition of fidelity has been given in section 1.3.2. A simulator fidelity typology has been proposed by Rehmann and his colleagues which includes the following three dimensions[28]:

- Equipment fidelity: The degree to which the simulator duplicates the appearance and feel of the real system.

- Environment fidelity: The extent to which the simulator duplicates motion cues, visual cues, and other sensory information from the task environment.
- Psychological fidelity: The degree to which the trainee perceives the simulation to be a believable surrogate for the trained task.

Miller on the other hand, made a distinction between psychological fidelity (see above) and engineering (or physical) fidelity with the latter described as 'the degree to which the training device or environment replicates the physical characteristics of the real task'[2]. Thus, in this distinction, engineering fidelity could be seen as a combination of equipment and environment fidelity. The fidelity of a simulation is never identical to the real thing due to cost, engineering technology limitations, avoidance of danger, ethics, psychometric requirements and time constraints[15].

The simulation experience must be real and motivating in order for participants to forget any less-than-perfectly-representative aspect of their task[2, 20]. Issenberg et al refer to fidelity as a form of simulator validity (face validity) and note that high simulator validity helps learners hone their visuo-spatial skills and responses to critical incidents[16]. The recall of the learning experience is better when the retrieval setting matches the setting of original learning (context dependence)[50]. However, the appropriate degree of fidelity depends on the desired outcome[47], as well as the type of task and stage of training[2]. Based on the activity theory (see section 2.3) fidelity is important for learning as it can help bring together the boundaries between simulation and 'real world' practice.

In order to achieve effective learning, educators should keep in mind the aforementioned characteristics and try to incorporate these when designing simulation training. Educational

design also needs to consider how interactions between different characteristics will influence the validity and effectiveness of the training. The following section is a brief description of current views on educational design in higher education.

2.5. EDUCATIONAL DESIGN

An important goal of medical education is for trainees to become experts. This requires trainees to attain a deep, organised and contextualised understanding of their discipline[51]. According to McGaghie et al mastery learning varies among learners and includes seven complimentary features[15]:

- Baseline testing
- A sequence of clear learning objectives in units of increasing difficulty
- Engagement in learning activities focused on reaching the objectives
- A passing standard for each learning unit (e.g. test score)
- Formative testing to measure unit completion at a preset minimum passing mastery standard
- Advancement to the next learning unit once the passing standard has been achieved
- Continued practice or study on a learning unit until the passing standard is achieved

The abovementioned features are in agreement with current research about effective learning and reflect tenets of good practice in terms of simulation course design. The quality of a simulation design is to a large extent responsible for its superiority, or lack of superiority, over other methods of training[20]. According to educational research, the design of higher

education learning environments should give emphasis to a constructivist pedagogical theory, a deep approach to learning (i.e. students should make sense of what is to be learned and develop their own perspectives and syntheses of the subject), a student-centred approach to teaching and outcomes-centred subject design[51, 52]. Deep learning is more likely when i) students' motivation is intrinsic, ii) learning activities are planned, active, reflected upon and processed, and related to abstract concepts, iii) learning includes interactions with others and iv) there is a well structured knowledge base, with new learning being related to existing knowledge rather than learned in isolation[52].

Based on the aforementioned approach, course design should include a) a definition of the course learning outcomes (i.e. what a student will be able to do at the end of the learning experience), b) assessment tasks for trainees to demonstrate that they can meet the course learning objectives, c) learning activities that provide trainees with opportunities to practice the aforementioned tasks and d) determining the knowledge base (content) needed for the learning activities to be carried out[51]. Identifying learning outcomes is usually done through an analysis of the topic (i.e. working out what the concepts to be learned are and what must be known before something else can be learned) or, in the case of tasks, through task analysis (i.e. working out the steps included in each task)[49]. Higher education principles provide a structured approach to course design and may be of help to educators designing a simulation course.

2.6. SIMULATION VALIDATION AND EFFECTIVENESS

Within the simulation educational context, *validity* measures whether the simulator actually is teaching what it is intended to teach[53] i.e. does the process of A in Fig. 2.1. There are many types of educational validity pertaining both to the delivery of learning materials and the measurement of educational outcomes[47]. The generalisability of simulation-based clinical learning to real patient care settings i.e. the transition from A to D in Fig. 2.1 has been referred to as concurrent validity[16]. Thus, simulation validation is inherently related to determining the effectiveness of simulation training through the causal chain described in section 2.2, which includes, but is not limited to, assessing validity. Validity is necessary but not sufficient for effectiveness. The most common quantitative study designs used in evaluating the effectiveness of simulation training, including the RCT design employed in this thesis, are presented in the following section.

2.6.1. STUDY DESIGNS FOR CONDUCTING EVALUATIONS OF SIMULATION TRAINING

This section provides an overview of the study designs used in conducting evaluations of simulation training. The quantitative study designs used in simulation training research include observational, quasi-experimental, and RCTs. These studies may be uncontrolled or controlled over time and/or space (i.e. may have a non-intervention group).

If studies do not have a non-intervention group, data collection needs to take place in the pre-intervention period, during the intervention and/or post-intervention. Observational before and after studies have a methodological weakness in terms of distinguishing cause and effect as any observed change could plausibly be attributed to developments other than the

intervention of interest[54]. In studies with a non-intervention group, comparisons can be made only after the training intervention or both before and after the intervention phase[54]. However, non-randomised postintervention comparisons are the least reliable as the intervention and control groups may have hidden inherent differences[54].

The RCT study design was chosen for the two trials of this thesis, which are reported in chapters 4 and 5. RCTs are often regarded as the gold standard with regards to evaluating the effectiveness of an intervention as they provide the best opportunity to demonstrate causality and to control bias[55, 56]. RCTs with both preintervention and postintervention measures are considered the strongest design[54] but such designs are not always feasible in the evaluation of simulation training. For example, preintervetion measures may not be possible for a completely new skill. The RCT reported in Chapter 4 includes baseline assessment on the simulator as *in vivo* assessment of skill prior to any training whatsoever would not be considered ethical. The RCT reported in Chapter 5 includes pre and post intervention measures. One note of caution is that while RCTs have been advocated as means of evaluating new training methods, they need to reflect normal educational practices[57].

Regardless of the study design employed to evaluate simulation training, there are a number of methodological considerations to be taken into account in all evaluations, which are presented in section 2.6.2. Qualitative methods should also be used to explore the ideas and concerns of simulation trainees. It has been argued that medical education is too complex, too local and too subjective an activity to be reduced to quantitative analysis only[58]. A growing body of medical education researchers view education and research in education as fundamentally humanistic endeavours and are choosing to adopt social science methodologies

which favour the local context[59]. Qualitative research can also help define the preliminary research questions which can then be addressed by quantitative studies and could also be used to inform theory and policy[60]. In addition, qualitative research can be used for triangulation of research outcomes.

2.6.2. ASSESSING CLINICAL SKILLS PERFORMANCE DURING EVALUATIONS OF SIMULATION TRAINING

A popular construct for the assessment of clinical skills performance is the Miller pyramid (Fig. 2.2), which includes four levels, with 'knows' and 'knows how' at the base (declarative knowledge) and 'shows how' and 'does' (procedural knowledge) at the top[16, 61-63]. Simulation can be used to assess the first three levels[16]. 'Does' can only be assessed through observing a practitioner working in the real world[61].

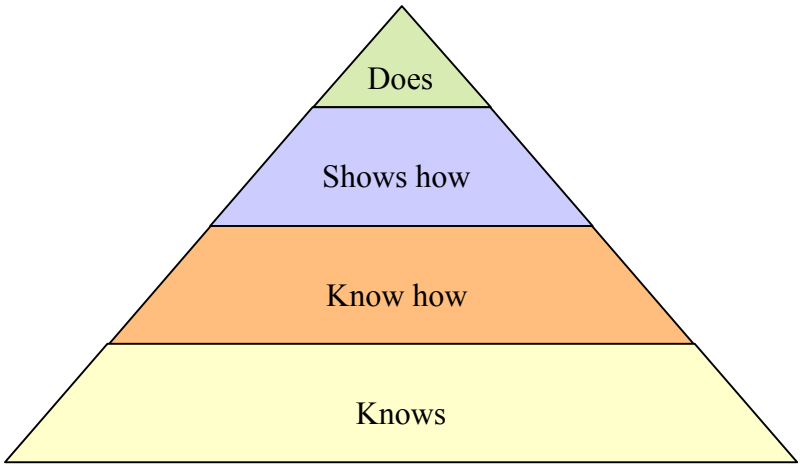

Figure 2.2. The Miller pyramid

When evaluating simulation training, the assessment should be suited to the aspects of performance being measured and should be valid, reliable and feasible[64, 65]. The determination of the validity and reliability of an assessment tool is an intricate process, as there are a variety of types of validity and reliability. Furthermore, there is considerable variation in the definitions of assessment validity found in the simulation training validation literature[66].

Test **validity** refers to the degree that a test is actually measuring what it was designed to measure[63]. The validity of a test can be described by:

- Content validity: 'the extent that the test measures all pertinent aspects of the competency being studied'[67]. To have a high content validity a test must cover the full range of the curriculum learning objectives[64].
- Construct validity: 'the extent to which the test fulfils expectations of differentiating the novice from the expert'[67], thus reflecting the construct that is being tested[64].
- Face validity: 'the extent to which the test reproduces what is experienced in real life'[67].
- Criterion validity: the extent that the test reflects the best existing 'gold standard' of practice[64, 67]. Criterion validity includes two subcategories; concurrent and predictive validity[66]. Concurrent validity is 'the degree to which scores on a test correlate with the scores on an established test'[65]. Predictive validity relates to 'the certainty with which a test can predict future performance' e.g. performance in the operating room[65].

The **reliability** of a test refers to the extent to which a test will produce a consistent measurement of the attribute or competency in question when applied under different occasions or by different observers. The reliability of a test may be measured by:

- Inter-rater reliability: the degree to which different raters of the test agree with each other[67].
- Intra-rater reliability: the intra-rater reliability of a test is estimated by presenting repeatedly the same observations to one rater[68].
- Internal consistency: the degree to which different parts of the test give consistent evaluations of the attribute or competency being tested[67].
- Test-retest reliability: The test-retest reliability of a test is estimated by presenting the same test to the same subjects two or more times[68].

The identification of reliable measures of learning can pose difficulties as the most relevant measures (e.g. surgical complications) may be too infrequent or unreliable, while alternative measures that may be more tractable to statistical analysis may be relatively poor measures of performance (e.g. performing an operation quicker does not necessarily indicate better performance as it may increase the risk of complications)[69].

The **feasibility** of a test should also be checked. The assessment should be feasible within the existing resources and time and the outcome must be achievable by the participant. Thus, the process of selecting assessment instruments should include a series of important considerations. These include the amount of time required to construct the instrument and conduct the marking process, the ease of interpretation of the results, the quality of feedback resulting from the instrument and the instrument covering important elements within the simulation course[65]. The assessment should also have a positive effect in terms of student motivation, good study habits and positive career aspirations i.e. the assessment should be acceptable to participants[65].

Multiple assessment instruments may be applied to a single simulation, which can be directed towards 'single' aspects of performance or can be an overall 'global' performance score[64]. Two general approaches to rating technical performance are global rating scales (subjective) and checklists (objective)[56] and these have been use in the RCTs reported in chapters 4 and 5. The types of assessment instruments used as measures of performance during existing simulation training evaluation studies can be seen in the summary tables provided in Chapter 3.

2.7. SUMMARY

This chapter has highlighted that a causal chain needs to be studied in order to determine the transfer of learning from simulation training to health outcomes. The link showing that improved training outcomes *in vitro* translate to improved clinical processes *in vivo* constitutes a very important step in this chain and is a key objective of simulation validation. The educational theories that can be used to explain how simulation training can indeed lead to improved performance have been described. Links have been outlined between these theories and the characteristics of simulation that facilitate learning i.e. feedback, repetitive practice, curriculum integration, range of training levels, multiple learning strategies, individualised learning, clinical variation, controlled non-patient-facing environment, defined outcomes and fidelity. Current higher education principles with regards to course design have been offered as a guide to designing simulation training. The evaluation of simulation training has been further discussed in terms of the study designs used with the validity, reliability and feasibility of the assessment tools being important regardless of the study design chosen.

CHAPTER 3

THE EFFICACY AND EFFECTIVENESS OF TECHNICAL AND NON-TECHNICAL SKILLS SIMULATION TRAINING: REVIEW OF REVIEWS

3.1. INTRODUCTION

The aims of this chapter are a) to identify and synthesise review-level material to highlight the current evidence on the efficacy and effectiveness of simulation training of technical and non-technical skills and b) to highlight conflicting evidence and gaps in the evidence linking simulation training to real-life clinical performance. A supplementary aim of this review is to investigate some of the conditions under which simulation training may be more favourable such as the type of skill being taught.

It would be ideal to undertake a comprehensive and systematic review of literature on the whole of the research involving simulation as an educational intervention. However, initial searches revealed this to be an unrealistic aim, given that for example, a scoping search of MEDLINE (Ovid) in April 2006 using the string 'simulation and (education or training or learning)' returned 4,951 hits. Therefore, the aim to provide an exhaustive bibliography of all existing literature related to educational uses of simulation was precluded by the timeframe of this research. On the other hand, an initial search of the MEDLINE, EMBASE, CINAHL, PsychInfo, Web of Science and the Cochrane database using the terms 'simulation', 'simulator', 'simulated' and 'systematic review' revealed a dearth of systematic reviews.

Thus, the scope of the search was extended to include reviews that used a less strict methodology.

3.2. METHODS

The following process was applied:

- Systematic searches of the literature on the MEDLINE and EMBASE databases using the search command *simulat$ AND (educat$ OR train$ OR learn$)* and limiting the results to review articles in English. The time limits were from 1995 to June 2009.

- Selection of relevant reviews: The titles of all citations returned by this search were screened to eliminate the obviously irrelevant ones. For the remaining, abstracts of the review articles were retrieved and stored in the Reference Manager 11.0 software and duplicates were removed. The remaining abstracts were screened according to the following a priori inclusion criteria:

  i) The articles are systematic reviews, literature reviews, syntheses or meta-analyses.

  ii) The articles review primary studies investigating the training efficacy and/or effectiveness of simulated patients, animal models, cadavers or a simulator device such as part-task trainers, manikins, computer based systems, virtual reality (VR) systems, integrated simulators and simulated environments.

  iii) The reviewed studies include a study population of medical, dental, nursing or allied health professionals/trainees/students/participants practicing a medical procedure e.g. cardiopulmonary resuscitation (CPR) using simulation.

  iv) The reviewed modes of simulation are used as a training intervention with measured learner outcomes.

v) The outcome measures in the reviewed studies are *in vitro* and/or *in vivo* practical procedures, clinical skills, patient investigation, patient management, health promotion, communication and/or decision making.

Primary studies, validation studies that did not include a relevant simulation training component, articles of purely descriptive nature, reviews that offered no new data, theory papers and opinion or position statements were excluded. For the abstracts identified, full-text articles were retrieved for review and were further considered against the inclusion criteria. Citation lists of relevant reviews were scanned in search of additional reviews. Two additional review articles were also included that had been previously identified through the references of a departmental report and through the initial scoping searches of the literature on simulation training in healthcare.

- Critical appraisal of the reviews that satisfied the aforementioned inclusion criteria was undertaken by adapting the Health Development Agency (HDA) evidence briefing approach and assessing the extent to which the reviews were[70]:
  - Systematic: 'does the review apply a consistent and comprehensive approach?'
  - Transparent: 'is the review clear about the processes involved?'
  - Analytically sound: 'are the appropriate methods of methodological analysis undertaken?'
  - Relevant: 'is the review relevant' to the population groups of medical, dental, nursing and allied health professionals/trainees/students that the review seeks to target?

  This process involved two stages (shown in Table 3.1). The first stage assessed the strengths of the methods (i.e. clear aim/research question, appropriate databases,

additional search strategies, specific inclusion criteria) used by each reviewer to identify and select the literature under review, while the second stage assessed the quality of its methodological analysis (rigour of individual studies assessed, individual studies' findings presented and analysed clearly and consistently) and the appropriateness of its conclusions[70].

To be included in the synthesis of evidence reported here, reviews had to have specified a clear aim or research question, identified appropriate source databases and specified their search terms and/or inclusion criteria. This relates to the reviews' level of being systematic and transparent on how individual studies were selected. Transparency is regarded as very important in order to obtain a balanced view of the evidence[70]. Reviews that did not satisfy this minimum critical appraisal threshold (Appendix 3.1) were not included in the synthesis of the evidence. The results of the critical appraisal of the reviews that passed this threshold are shown in Table 3.1. Great effort was made for this process to be as objective as possible. However, the author acknowledges that decisions do contain a subjective element, which should be taken into account when reading the findings[70].

- Categorisation of the evidence as Core or Supplementary: The reviews that were judged to pass the aforementioned minimum critical appraisal threshold were categorised as 'Core' or 'Supplementary'. This categorisation was made based on how closely their aim matched this review's aim. Papers that specifically addressed a research question on whether simulation is an effective method of training for clinical skills were

classified as Core reviews. Supplementary review papers only addressed this question indirectly e.g. as part of a wider research question or added further insight on when simulation training is effective. Thus, despite not being closely matched to this review's aim, these reviews were judged to have something useful to say about the nature of the evidence, adding further insights on the efficacy and/or effectiveness of the interventions under review[70].

- Data extraction: Where a review passed the aforementioned minimum critical appraisal threshold, information was extracted, where available, on the relevant individual studies of that review in terms of the study design, the procedure being assessed, the population, comparators, outcome measures, blinding and results. Data were extracted directly to tables on a Word document as seen on Appendices 3.2 and 3.3. Self-report data e.g. participant self- assessments were not included.

- Synthesis of the evidence for the different modes of simulation training: A synthesis of the findings of individual studies that were extracted from the Core and Supplementary reviews is presented in section 3.3 using where possible the taxonomy shown in Fig. 3.1 below.

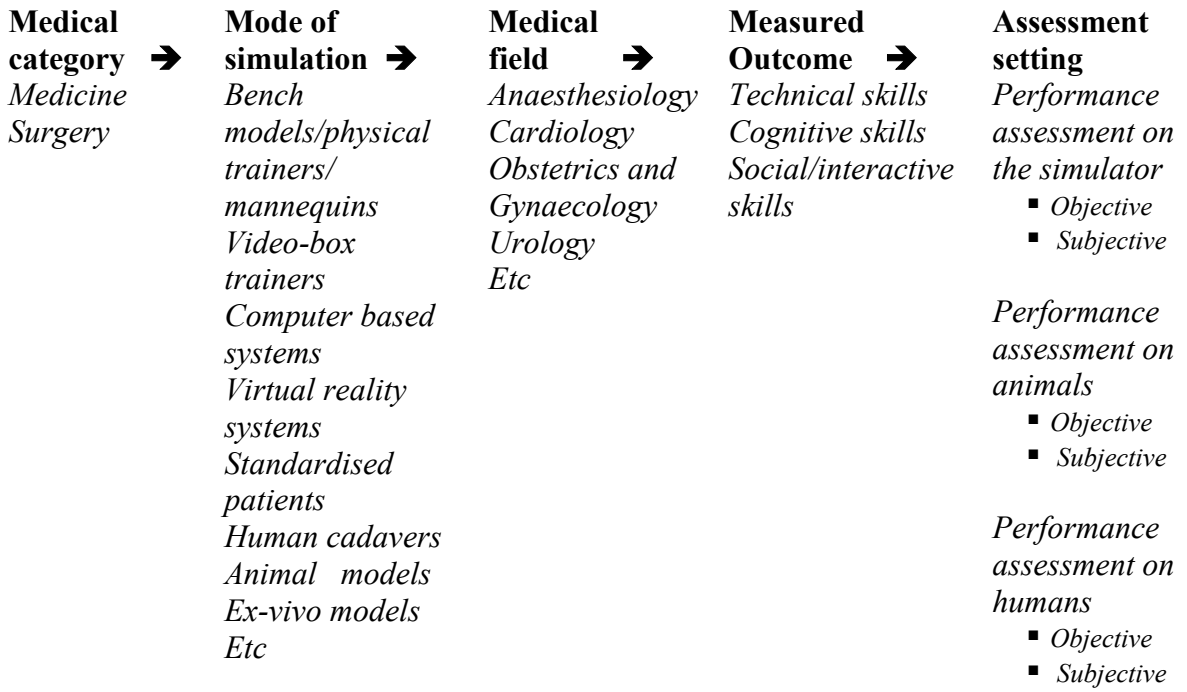

Figure 3.1. Taxonomy of different modes of simulation training

The synthesised findings were exclusively obtained from the corresponding reviews without referring back to the individual studies. The extracted review data of the individual studies were categorised in terms of medical category, mode of simulation and assessment settings (*in vivo* or *in vitro*) and synthesized in terms of comparators and results. However, it was not always possible to further define comparators classified as 'standard training' as further information was not always provided in the reviews. Individual study findings were identified as positive (√) if simulation training had been reported to have resulted in better outcomes than the comparators and as negative (X) if simulation training had been reported to have resulted in worse outcomes than the comparators. Some reviewers did not always state explicitly whether these differences were statistically significant. A 'no difference' (-) finding was recorded if simulation training and its comparators had resulted in similar outcomes. For each mode of simulation the findings of all individual studies were summarised into a single

verdict regarding the overall direction of the review evidence on that mode *versus* each comparator. Where individual study results were conflicting, the overall evidence was identified as inconclusive (?). These verdicts were qualitative with each individual study given equal weight. Due to variability in the individual study comparators and reported outcome measures meta-analysis was not attempted.

3.3. RESULTS

The MEDLINE search returned 901 hits while the EMBASE search returned 1160 hits. All citations were screened and after eliminating the obviously irrelevant ones and removing the duplicates, 382 citations were retained and their abstracts were read and checked against the inclusion criteria. Eighty-two citations were further retained and their full-text articles were retrieved for review and further consideration against the inclusion criteria. An additional review[36] and a meta-analysis[71] were also considered. The latter two papers had been previously identified from the references of an unpublished departmental report[72] and through a previous scoping MEDLINE search. The full text review led to the further exclusion of 46 articles. All 38 articles underwent critical appraisal. A flowchart of the review selection process is presented in Fig. 3.2.

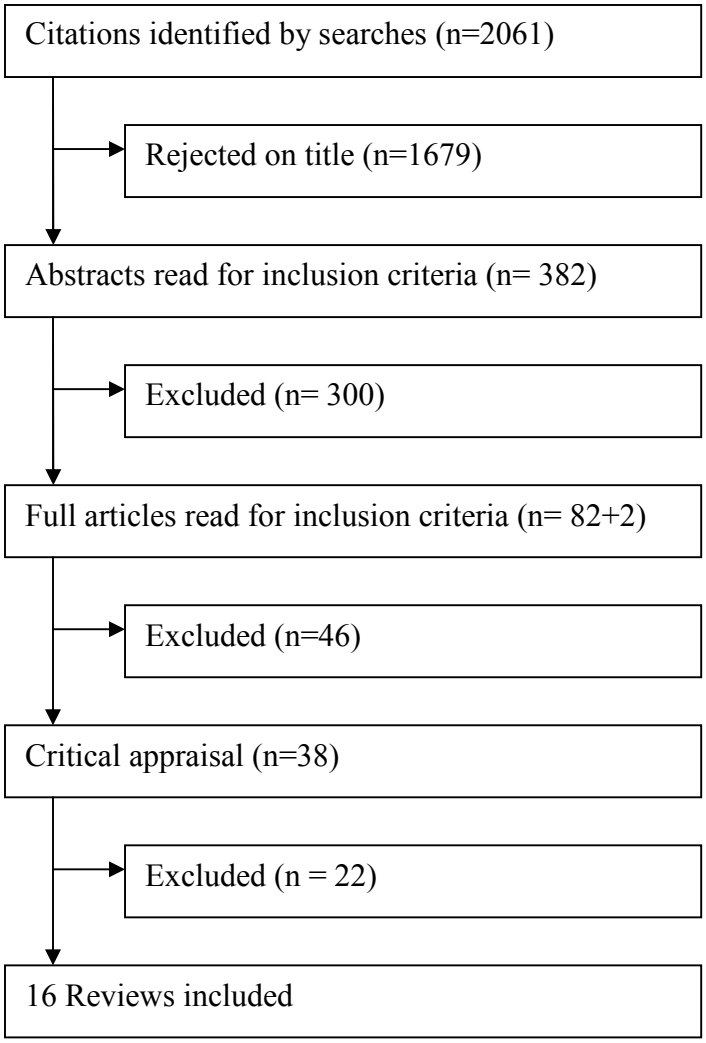

Figure 3.2. Flowchart of the review process

The results of the critical appraisal can be found in Table 3.1 (included reviews) and Table 3.2 (excluded reviews) (Appendix 3.1). Of the 38 reviews included in the tables, only 16 were judged to pass the adopted minimum critical appraisal threshold and were included in the final synthesis of the evidence. These included ten Core and six Supplementary review papers (Table 3.1). Of the 16 reviews and meta-analyses included eight focused on surgery[8, 21, 71, 73-77], four focused on medicine[78-81] and four pertained to both fields[16, 82-84].

Table 3.1. Critical appraisal of the reviews using a summary of the HDA's critical appraisal tool (√ = Yes, X = No/Not reported, - = the review format precludes judgment)

	Stage one					Stage two				
Author and date	Specifies clear aim or research question	Identifies appropriate range of source databases	Undertakes additional search strategies*	Specifies search terms	Specifies inclusion criteria	Rigour of individual studies assessed	Individual studies' findings presented clearly and consistently	Individual studies' findings analysed clearly and consistently	Conclusions presented relate to individual studies' findings	**Review category**
Arnold & Farrell 2002[73]	√	√	√	X	√	√ (in part)**	√	X	√	*Supplementary*
Aucar et al 2005[74]	√	√	√	√	√	X	X	X	√	*Supplementary*
Byrne et al 2008[82]	√	√	√	√	√	√ (in part)**	√	√	√	*Core*
Gaffan et al 2006[78]	√	√	√	√	√	X	√	√	√	*Supplementary*
Gerson 2006[21]	√	√	X	√	X	√	√	√	√	*Supplementary*
Gurusamy et al 2008[75]	√	√	√	√	√	√	√	√	√	*Core*
Hamilton 2005[79]	√	√	√	√	√	X	X	X	–	*Supplementary*

*Additional search strategies involve follow-up of references/journals, consultation with experts in the field and searching for grey literature
**i.e. rigour was addressed only for some of the studies described or for a specific aspect of the studies

Table 3.1. Critical appraisal of the reviews using a summary of the HDA's critical appraisal tool (continued)

	Stage one					Stage two				
Author and date	Specifies clear aim or research question	Identifies appropriate range of source databases	Undertakes additional search strategies*	Specifies search terms	Specifies inclusion criteria	Rigour of individual studies assessed	Individual studies' findings presented clearly and consistently	Individual studies' findings analysed clearly and consistently	Conclusions presented relate to individual studies' findings	**Review category**
Haque & Srinivasan 2006[71]	√	√	√	√	√	X	√	√	√	*Core*
Issenberg et al 2005[16]	√	√	√	√	√	√	X	√	√	*Supplementary*
Lane & Rollnick 2007[80]	√	√	X	√	√	√	X	X	√	*Core*
Lynagh et al 2007[83]	√	√	√	√	√	√	√	√	√	*Core*
McGaghie et al 2006[84]	√	√	X	X	√	X	√	√	√	*Supplementary*
Ravert P 2002[81]	√	√	√	√	√	√ (in part)**	√	√	√	*Core*
Sturm et al 2008[76]	√	√	√	√	√	√	√	√	√	*Core*
Sutherland et al 2006[77]	√	√	√	√	√	√	√	√	√	*Core*
Tsang et al 2008[8]	√	√	√	√	X	X	√	X	√	*Supplementary*

*Additional search strategies involve follow-up of references/journals, consultation with experts in the field and searching for grey literature
**i.e. rigour was addressed only for some of the studies described or for a specific aspect of the studies

The findings of the Core and Supplementary reviews as reported by the reviewers are described in detail below. Evidence reported in length mainly originates from Core reviews unless otherwise indicated. However, in the absence of Core reviews, Supplementary reviews have been used to provide an indication of the existing research. Component studies are summarised in Tables 3.3-3.16 and in Tables 3.18-3.25 (Appendices 3.2. and 3.3) and have been synthesised according to the first two levels of the taxonomy presented in Fig. 3.1 (Surgery/Medicine and mode of simulation), being classified where possible according to the mode of simulation employed. The results of their overall synthesis are presented in Tables 3.17 and 3.26.

3.3.1. SIMULATION TRAINING IN SURGERY

Six Core papers - five systematic reviews[75-77, 82, 83] and a meta-analysis[71] - met the HDA minimum appraisal criteria (Table 3.1). These are:

- Byrne A.J., Pugsley,L., Hashem,M.A. Review of comparative studies of clinical skills training. Medical Teacher 2008; 30:764-767
- Haque, S., Srinivasan, S. A meta-analysis of the training effectiveness of virtual reality surgical simulators. IEEE transactions on information technology in biomedicine 2006;10(1):51-58
- Lynagh, M., Burton, R., Sanson-Fisher, R. A systematic review of medical skills laboratory training: where to from here? Medical Education 2007; 41: 879-887
- Gurusamy, K., Aggarwal, R., Palanivelu, L., Davidson, B.R. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. British Journal of Surgery 2008; 95: 1088-1097
- Sturm, L.P., Windsor, J.A., Cosman, P.H., Cregan, P., Hewett P.J., Maddern, G.J. A systematic review of skills transfer after surgical simulation training. Annals of Surgery 2008; 248 (2): 166-179
- Sutherland, L., Middleton, P., Anthony, A., Hamdorf, J., Cregan, P., Scott, D., Maddern, G. Surgical Simulation. A Systematic Review. Annals of surgery 2006; 243(3):291-300

A total of 42 primary studies were covered, 26 (62.9%) of which appeared in at least two of these papers. All were RCTs of a training technique using at least some elements of simulation compared with another method of surgical training, or no surgical training. However, Haque & Srinivasan and Gurusamy et al focused on VR simulators.

Byrne et al[82] reviewed studies that compared methods used to train staff in the clinical skills of venous cannulation, intubation and central venous line insertion. Although they did extract data pertaining to the quality of the studies including randomisation and whether the assessment tool used was validated, they did not report on whether assessors were blinded. They included nine studies pertaining to some form of simulation training; eight were prospective and one was retrospective. They reported that the studies in their review had methodological defects including small numbers of participants and lack of validated assessment tools with the exception of three cases. Sample sizes ranged from 26 to 163 participants with a total of 855. This was the only core review that specifically addressed low - fidelity mannequin training.

Gurusamy et al[75] reviewed 23 RCTs that evaluated the effectiveness of VR training with the aim of determining whether VR training could 'supplement and/or replace conventional laparoscopic training' for surgical trainees. This was a Cochrane review. They assessed the methodological quality of their studies in terms of randomisation, generation of allocation sequence, allocation concealment, blinding and follow-up. Fifteen of these were judged to have sufficient blinding but overall only three trials were judged to have a low risk of bias. Sample sizes ranged from 10 to 65 with a total of 612. This was a high quality review; however the sample sizes of the included studies were small.

Lynagh et al[83] reviewed the effectiveness of medical skills simulators and laboratories and reported on 37 RCTs pertaining to surgical procedures. They report that 19 of these were blinded, 3 were not and 15 did not state whether blinding had been used. The reviewers also reported that 20 of these studies had investigated skills transfer by outcome measures *in vivo,* on patients or live animals. However, they also noted that several of the studies had limitations in terms of small sample sizes, lack of reporting on randomisation methods, non-standardisation of the training conditions and non-validated outcome measures. Sample sizes ranged from 6 to 163 participants with a total of 1005. This was a high-quality systematic review; the authors reported consistently on the methodological limitations of the included studies and were the first to investigate systematically the issue concerning whether skills acquired through simulator training are retained over time.

Sturm et al[76] reviewed the surgical skills transfer to the real-life operative setting following simulation-based training. Their review included 10 RCTs and one non-randomised comparative study. The reviewers reported that three of these studies were not blinded and that more than half did not provide sufficient methodological details such as exclusion criteria, randomisation, allocation concealment, study period and intention to treat analysis. Again, sample sizes were small ranging from 8 to 45 participants with a total of 238. This was one of the few systematic reviews that specifically addressed the transfer of skills to the real-life setting following simulation training. However, the small number of available studies, their variable quality, the small sample sizes and the lack of uniformity in measurement of outcomes limits the strength of the conclusions. Furthermore, this review was confined to laparoscopic cholecystectomy and colonoscopy/sigmoidoscopy, which may limit the applicability of the findings to other fields of simulation training.

Sutherland et al[77] have also reviewed the instructional effectiveness of surgical simulation. In terms of quality, they report that only three out of the 30 studies in their review were likely to have had adequate allocation concealment and only half had reported using blinded assessors with one additional study having one blinded assessor out of the two. The rest of the studies did not did not state whether their outcome assessors were blinded[77]. The same reviewers also report that 'previous experience of participants varied greatly between the studies' and so did the length of time devoted to training. The shortest training period reported was 10 minutes and the longest was 10 hours while some other studies reported 'number of sessions or the number of tasks that were practiced'. In addition, the reviewers note that the large number of comparisons made within and between studies could have diluted the power to detect differences and that comparisons may have been confounded by factors such as mentoring. However, most studies did not have losses to follow-up[77]. Sample sizes were small, ranging from 10 to 49 participants with a total of 760. Again, the variable quality of the studies, the small sample sizes and the lack of uniformity in measurement of outcomes limits the strength of the conclusions. Furthermore, the authors use the generic term 'computer simulation' instead of the more specific term 'VR simulation'.

Haque & Srinivasan[71] conducted a meta-analysis of the efficacy of VR training in achieving skills transfer from the simulator to the operating room. The meta-analysis included only seven prospective studies as inclusion was based on the availability of specific statistically relevant data (task completion time and error score). The authors did not provide detailed information on the rigour of these studies. They only report that these studies were 'on the whole' sufficiently randomised and that 'both groups were reported to be equally qualified

and experienced' in terms of surgical skills. Sample sizes were small, ranging from 6 to 29 participants with a total of 127.

Four Supplementary reviews met the minimum critical appraisal criteria and are also shown in Table 3.1. These are:

- Arnold, P., Farrell, M. Can virtual reality be used to measure and train surgical skills? Ergonomics 2002;45(5):362-379
- Aucar, J., Groch, N., Troxel, S., Eubanks, S. A review of simulation with attention to validation methodology. Surg Laparosc Endosc Percutan Tech 2005; 15(2): 82-89
- Gerson, L.B. Evidence-based assessment of endoscopic simulators for training. Gastrointest Endosc Clin N Am 2006; 16: 489-509
- Tsang, J.S., Naughton, P.A., Leong, S., Hill, A.D.K., Kelly, C.J., Leahy, A.L. Virtual reality simulation in endovascular surgical training. Surgeon Journal of the Royal College of Surgeons of Edinburg & Ireland 2008; 6(4):214-20

Aucar et al[74] reviewed the 'state of the art' regarding the development and use of simulation technology for the training and assessment of surgical technical skills. Their main focus was the validation methodology of simulation studies. They included 23 studies on both surgical assessment and training rather than focussing on training alone. The authors did not provide information on the rigour of individual studies reviewed, included studies of various study designs and did not attempt to synthesise studies' findings in a consistent manner. The review included mechanical, electronic or combined models pertaining to open or laparoscopic surgery but did not make a clear distinction among these when reporting study findings. The majority of the studies reviewed concerned laparoscopic surgical techniques with 12 of them pertaining to training or assessment on the MIST-VR simulator. The studies reviewed by Aucar et al[74] regarding training can also be found in the Sutherland et al[77] review with the exception of two studies.

Arnold & Farrell[73] also addressed the validation and efficacy/effectiveness of simulators for training without making a distinction between the studies addressing these two purposes. They provided information on three additional studies pertaining to VR training in palpation for tumor detection, catheter placement and laser coagulation. Their review contains studies only up to 2000 so it could be considered outdated given that the field of VR technology is rapidly evolving. Since then, numerous studies have been published on the efficacy and effectiveness of VR simulators, which can be found in the Core reviews of this chapter.

Gerson[21] reviewed the evidence regarding the current generation of endoscopic simulators and their efficacy in training. This review appears comprehensive on the whole; the author reports that it is the product of a systematic review and included research questions on the efficacy of endoscopic simulators. However, this is only one of several research questions and they do not address each of their questions separately but discuss validation and training intervention studies simultaneously. Thus, this paper was not classed as a Core review. Further, the author did not specify the inclusion criteria used. Gerson[21] provided evidence from 11 studies pertaining to VR and animal-based simulators; four of these were randomised trials while the rest were cohort or case-control studies. Blinded assessors were reported for only two studies.

Tsang et al[8] reviewed the potential benefits of the use of VR simulation in vascular surgery training. Their review included five prospective studies of pre-test post-test design, which did include a comparator, and two RCTs. None of these studies had been previously covered by the other reviews included in this review. However, this review was classed as supplementary as the authors did not specify their inclusion criteria and did not address any aspects of the

rigour of the included studies such as blinding. Sample sizes ranged from 9 to 29 participants with a total of 139.

3.3.1.1. MODELS OR PHYSICAL TRAINERS

The Core reviews by Sutherland et al[77] and Lynagh et al[83] identified seven RCTs pertaining to model simulation including one RCT pertaining to physical trainer simulation. All studies used a surgical performance outcome measure. However, the specific technical skills measured are not described in the reviews. Two of the studies used checklists and global ratings and one used time to complete the task, but the method of assessing performance is not specified for the remaining studies. Thus, it is not clear whether performance assessments were objective or subjective. The results of the studies reviewed and the outcome setting are presented in Tables 3.3 – 3.5 (Appendix 3.2).

Physical trainer/model training vs no training

With regards to model training compared to no training, the three RCTs reviewed have led to inconsistent results *in vivo*. Simulation training was reported to be superior to no training in 5 out of 8 assessment areas of laparoscopic hernia repair in the operating room. Two further studies of laparoscopic performance on an anaesthetised pig led to inconsistent results, with one study reporting improved performance for the intervention group while the other study reported no difference (Table 3.3).

Physical trainer/model training vs other forms of training

With regards to model training compared to other forms of training, the evidence *in vitro* and *in vivo* appears to be favourable. Out of the RCTs reviewed, 4 favoured simulation when compared to bedside training, learning from a manual and didactic training (2 studies) (Table 3.4). Furthermore, two of the studies found no differences between bench model training and video-box training and between bench model training and cadaver training (Table 3.5). Finally, one study reported inconsistent results with only 3 out of 7 laparoscopic outcomes favouring VR simulation over Physical Trainer training.

3.3.1.2. VIDEO-BOX TRAINERS

The Core reviews by Sutherland et al[77] and Lynagh et al[83] identified 20 RCTs pertaining to video-box trainer simulation. The studies used different technical skills as their outcome measures, including total performance scores, economy of movement, time, errors, respect of tissue, instrument handling, use of assistants, procedure knowledge, motion, flow of operation, number of finished stitches, needle placement accuracy and suture strength. However, outcome measures are not described further so it is not always clear whether assessments were objective or subjective. The results of the reviewed studies and the outcome setting are presented in Tables 3.6- 3.10 (Appendix 3.2).

Video-box training vs no training

Comparing video-box training to no training has led to positive results *in vitro* (Table 3.6). Out of the four RCTs reviewed, three favored simulation training over no training in terms of surgical performance and time taken assessed on the box-trainer (2 studies) or the MIST-VR

(1 study). The remaining study led to inconsistent results while assessing clip application on a water-filled glove with the intervention group showing superior economy of movements (EOM) but similar times and numbers of errors compared to the control group.

However, comparing video-box training to no training has led to inconsistent results *in vivo* (Table 3.6). Out of the four RCTs reviewed, one found superior laparoscopic performance on a pig in favour of the intervention group while the other found no difference. The remaining two studies, which are by the same researchers and may actually overlap, led to inconsistent results with the intervention group being superior in 3 out of 7 laparoscopic cholecystectomy outcomes in the operating room (OR) and similar to the controls for the remaining outcomes.

Video-box training vs standard training

Comparing video-box training to standard training (didactic instruction) did not lead to statistically significant differences in terms of times and performance *in vitro* in the one RCT found in the reviews (Table 3.7). Video-box training produced similar results to didactic instruction when endourological skills were assessed on a video-box trainer.

Video-box training vs simplified simulation training

Comparing video-box training to simplified simulation (simplified mirrored box) training did not lead to statistically significant differences in terms of laparoscopic skills performance in one RCT (Table 3.8). Performance was measured on both the video-box and the mirrored box.

Video-box training vs VR training

Comparing video-box training to VR training has led to inconsistent results *in vitro* (Table 3.9). Out of the eight RCTs encountered, four found no statistically significant differences in terms of times (4 studies) and surgical performance (1 study). Conversely, two of the RCTs found that the VR training lead to superior laparoscopic performance and one had inconsistent results.

Among the reviewed studies, only two RCTs were found to have compared box-training to VR training *in vivo*. One study found that VR training was superior in terms of laparoscopic cholecystectomy performance in the OR. The other found no differences in terms of time and errors while evaluating laparoscopic skills on a pig.

Comparing different types of video-box training

Two RCTs compared *in vitro* video-box training to extended video-box training with additional instruction leading to inconclusive results (Table 3.10). Both RCTs studied laparoscopic skills training and while one of them found no differences in participants' performance the other led to inconsistent results with the extended training group being superior in terms of object passing errors but no different in terms of other tasks.

3.3.1.3. VR SIMULATORS

The Core reviews by Byrne et al[82], Gurusamy et al[75], Sturm et al[76], Sutherland et al[77], Lynagh et al[83] and Haque and Srinivasan[71] identified 37 RCTs pertaining to VR simulation. Among these, 12 (32.4%) used a MIST-VR simulator. The reviewed studies used different technical skills as their outcome measures. Outcome measures included total performance scores, time, economy of movement, errors, accuracy and correct incisions. However, outcome measures are not always described and so it is not always clear whether assessments were objective or subjective. The Supplementary reviews of Gerson[21], Aucar et al[74] and Arnold & Farrell[73] and Tsang[8] identified 18 additional studies including four RCTs, two clinical trials, nine prospective studies and three studies of unspecified design. The results of the reviewed studies and the outcome setting are presented in Tables 3.11- 3.14 (Appendix 3.2).

VR training studies using a pre-test post-test design

Out of the nine prospective studies covered in the supplementary reviews, seven studies looked at the efficacy of VR training *in vitro* and two *in vivo* leading to inconsistent results (Table 3.11). Two of the studies investigated colonoscopy skills, six studies investigated vascular skills and one study investigated IV insertion. In terms of colonoscopy, participants in the first trial achieved improved examination efficiency on the simulator over time but this did not correlate to bed-side training cases. In addition, colonoscopy performance in the second study did not improve over five attempts in the absence of feedback. In the vascular studies *in vitro*, VR training led to improved procedure time (5 studies), contrast volume (4 out of 5 studies) and improved suturing performance in vascular anastomosis (1 study). However study results were inconsistent in terms of fluoroscopy times (5 studies). Finally, no

statistically significant differences were found in terms of IV insertion success rate and performance in humans.

VR training vs no training

Comparing VR training to no training has generally led to positive results *in vitro* (Table 3.12). Out of the thirteen RCTs that were reviewed, VR training lead to superior performance in eleven. Outcome measures included times, number of movements, overall performance and number of correct incisions. Two RCTs led to inconsistent results with VR training found superior in some outcomes (economy of movements, speed) but similar to no training in others (time taken, errors, laparoscopic tasks).

In addition, comparing VR training to no training has generally led to positive results *in vivo* (Table 3.12). Out of the seventeen RCTs that were reviewed, VR training led to superior performance in pigs and humans in twelve in terms of navigation speed, accuracy, times and number of complete procedures. Four studies led to inconsistent results with VR training leading to improved error scores/accuracy (3 out of 4 studies) but not better times. Thus, results regarding whether VR training leads to improve times *in vivo* appear inconsistent. Finally, one RCT found no difference in performance scores in terms of laparoscopic appendectomy in a pig.

VR training vs standard training

Comparing VR training to standard training has led to inconclusive results *in vitro* (Table 3.13). Out of the two RCTs that were found, VR training led to similar times when compared to didactic instruction. The remaining RCT compared two VR simulators to plastic arm training and *in vivo* practice and found that VR training led to superior IV cannulation performance in a simulated arm. However, this was the case only for one of the VR simulators. The reviewers did not make clear the standard training comparators that this result applied to and did not provide information on how the other VR simulator compared to standard training.

Comparing VR training to standard training has also led to inconclusive results *in vivo* (Table 3.13). Out of the nine RCTs that were reviewed, VR training led to superior performance on patients in four studies, inconsistent results in one study, inferior outcomes in three studies and absence of statistically significant differences in one study. In three of these studies, the intervention group had received the VR training in addition to the standard training given to the control group. Outcomes that were improved with simulation training included endoscopy performance, error scores, task completion time and number of errors and less patient discomfort. The comparators were standard training using mannequins (2 studies), bedside training (five studies), traditional fiber optic intubation (FOI) training (not further defined in the reviews, 1 study) and standard laparoscopic cholecystectomy training (not further defined in the reviews, 1 study). However, in three RCTs VR training was found inferior to traditional training in terms of IV cannulation success, IV catheter placement knowledge and flexible sigmoidoscopy performance and in humans. In one study the addition of VR training to standard training led to less patient discomfort but no significant difference

in flexible sigmoidoscopy performance. An additional study of unspecified design found no difference in retinal photocoagulation efficiency in patients.

Comparisons of different types of VR training

Three RCTs and one study of unspecified design were reviewed, which compared different types of VR simulation *in vitro* (Table 3.14). The first RCT found that medium level VR training was more effective than easy VR training in terms of surgical performance scores in the same MIST-VR simulator. However, the second RCT found that easy level VR training in clip application was better than difficult level VR training in terms of speed and blood loss tested at the same level of training. In the third RCT, 20 minutes of MIST-VR practice distributed in five minute blocks led to better laparoscopic performance scores in the MIST-VR than massed practice and 15 minutes of distributed practice. The final study compared five minutes of VR training to a control group who had only had 1.5 minutes on a liver tumor palpation VR simulator leading to no significant differences in terms of search time, location and differentiation of the tumor.

One RCT compared basic VR training in six tasks to basic VR training plus additional knot tying VR training *in vivo* on an anaesthetised pig. The extended training led to a better objective error score than the basic training for tying a surgical knot but no differences in subjective error scores and times for driving a needle through tissue and tying a knot.

3.3.1.4. EX-VIVO MODELS

Ex-vivo models have been used for endoscopic training and involve the placement of animal (porcine) gastrointestinal organ packages into a plastic human model[85-87]. The Core review by Lynagh et al[83] identified one RCT pertaining to ex-vivo endoscopic simulators for endoscopic haemostasis skills, which was also reviewed by Gerson[21]. The latter reviewer also identified a prospective study on the same simulator (Table 3.15 – Appendix 3.2). Both studies favoured the use of this endoscopic simulator (CompactEASIE) for haemostasis skills training.

The RCT was partly blinded and compared a mix of ex-vivo simulation and bedside training to bedside training. The simulator group improved significantly in terms of all the haemostasis skills investigated (precision using a coagulator, variceal ligation, injection and coagulation, hemoclip application), while the bedside group improved only in variceal ligation[21].

The prospective study looked at performance on the simulator following animal-based simulation training without using any comparators. Training with the ex-vivo model led to improvement in endoscopic haemostasis (subjective expert grading) for all participants with experienced participants improving more than participants with no experience. However, the assessors in this study were not blinded.

3.3.1.5. CADAVERS

The Core reviews by Lynagh et al[83] and Sutherland et al[77] identified one RCT comparing cadaver training to two different forms of training *in vitro* on a cadaver (Table 3.5, Table 3.16 – Appendix 3.2). This was a blinded study that compared cadaver training versus model simulator training versus standard learning from manuals. Performance was assessed on six surgical tasks in cadavers. There were no significant differences reported between model and cadaver training. The cadaver trained group received better global and checklist scores than the standard training group. However, Sutherland et al[77] noted that the researcher did not state whether these differences were statistically significant.

An additional study (Table 3.16) compared central venous line insertion cadaver training to no training *in vivo*. This was a retrospective study with historical controls where cadaver training was found to be better than no training in terms of the rate of pneumothorax in patients.

3.3.1.6. OVERALL FINDINGS REGARDING SIMULATION TRAINING IN SURGERY

The overall conclusions resulting from the synthesis of the individual studies, which were extracted from the reviews, can be seen in Table 3.17.

Table 3.17. Overall summary of the review evidence on simulation vs other comparators pertaining to surgery

Intervention	Author and Date	Comparator				
		No training	Standard Training	Model training/ Physical trainer	Video-box training	Cadaver training
Model training	Lynagh et al 2007[83], Sutherland et al 2006[77], Aucar et al 2005[74], Byrne et al 2008[82]	*In vivo*?	*In vitro*√ *In vivo*√	N.E.	*In vitro* -	*In vitro* -
Video-box training	Lynagh et al 2007[83], Sutherland et al 2006[77], Sturm et al 2008[76]	*In vitro*√ *In vivo*?	*In vitro* -	*In vitro* -	N.E.	N.E.
VR simulators	Lynagh et al 2007[83], Sutherland et al 2006[77], Haque & Srinivasan 2006[71], Gerson 2006[21], Aucar et al 2005[74], Arnold & Farrell 2002[73], Gurusamy et al 2008[75], Sturm et al 2008[76], Tsang et al 2008[8], Byrne et al 2008[82]	*In vitro*√ *In vivo*√	*In vitro*? *In vivo*?	*In vivo*?	*In vitro*? *In vivo*?	N.E.
Ex-vivo model training	Lynagh et al 2007[83], Gerson 2006[21]	N.E.	*In vitro*√ *In vivo*√	N.E.	N.E.	N.E.
Cadaver training	Byrne et al 2008[82], Lynagh et al 2007[83], Sutherland et al 2006[77]	*In vivo*√	*In vitro*√	*In vitro* -	N.E.	N.E.

(√ = Intervention superior, X = Intervention worse, ?= inconclusive evidence, - = no difference, N.E.=no systematically reviewed evidence found)

Model training

- Model training vs no training (3 studies): Inconsistent results *in vivo*
- Model training vs standard training (4 studies): Favourable results *in vitro* and *in vivo*

These findings appear contradictory and may be due to the specific model trainers used and different outcomes assessed. It appears that further research is needed to confirm the effectiveness of model training.

- Model training vs video-box training (1 study): No statistically significant differences *in vitro*
- Model training vs cadaver training (1 study): No statistically significant differences *in vitro*
- Physical Trainer training vs VR training (1 study): Inconsistent results *in vivo*

These studies raise a question with regards to the skills where a simple model may be as effective as higher cost simulators. It may be that the psychomotor demands of the task being practiced (e.g. lack of touch sensation during laparoscopic surgery as opposed to open surgery) can be met sufficiently by a simple model. However, these findings are based in isolated studies and hence, further research is needed.

Video - Box trainers

- Video-box training vs no training (7 studies): Favourable results *in vitro*, inconsistent results *in vivo*
- Video-box training vs standard training (1 study): No statistically significant differences *in vitro*

- Video-box training vs simplified simulation training (1 study): No statistically significant differences *in vitro*

Further research is needed to investigate the effectiveness of video-box training.

- Video-box training vs extended video-box training (2 studies): Inconclusive results *in vitro*

Again, it appears that further research is needed to investigate the optimum amount of video-box training.

Video-Box trainers vs VR simulators

- Video-box training vs VR training (7 studies): Inconsistent results *in vitro*
- Video-box training vs VR training (2 studies): Inconsistent results *in vivo*

Further research is needed to determine which of the two methods of simulation leads to better training outcomes.

VR simulators

- VR training without a comparator (9 studies): Inconsistent results *in vitro* and *in vivo*. Results vary depending on the skill tested with some favourable results for vascular skills.
- VR training vs no training (16 studies): Favourable results *in vitro* and *in vivo*
- VR training vs standard training (12 studies): Inconclusive results *in vitro* and *in vivo*

Further research is needed to compare VR training to standard surgical training methods.

- Medium level VR training vs easy VR training (1 study): Favourable results *in vitro* for the medium level VR training
- Easy level VR training vs difficult VR training (1 study): Favourable results *in vitro* for the easy level VR training
- Distributed VR training vs massed VR training (1 study): Favourable results *in vitro* for the distributed practice
- Basic VR training VS extended knot tying VR training (1 study): Inconsistent results *in vivo*

Further research may be useful to confirm the optimum format and difficulty level when practicing different surgical skills on a VR simulator.

Ex-vivo model training

- Ex-vivo model training without a comparator (1 study): Favourable results *in vitro*
- Ex-vivo model training plus standard training vs standard training (1 study): Favourable results *in vitro* and *in vivo* for the simulation plus standard training group

This research suggests that ex-vivo training may be of benefit in haemostasis training but further research is needed to confirm the effectiveness of ex-vivo simulators for haemostasis and explore its training potential regarding other surgical skills.

Cadavers

- Cadaver training vs no training (1 study): Favourable results *in vivo*
- Cadaver training vs standard learning from manuals (1 study): Favourable results *in vitro* for the cadaver group

- Cadaver training vs model training (1 study) : No statistically significant differences *in vitro*

Further research may be useful to verify these findings, including *in vivo* randomised controlled studies.

3.3.2. SIMULATION TRAINING IN MEDICINE

In addition to Lynagh et al[83] and Byrne et al[82] who also covered surgery, two Core reviews by Lane & Rollnick[80] and by Ravert[81] met the minimum appraisal criteria (Table 3.1) and are relevant to the efficacy and effectiveness of simulation training in medicine:

- Lane, C., Rollnick, S. The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. Patient Education and Counseling 2007; 67: 13-20
- Ravert, P. An integrative review of computer-based simulation in the education process. Computers, Informatics, Nursing 2002;20(5): 203-208

Unlike Lynagh et al[83], these two reviews included less strict inclusion criteria in terms of study design and did not limit their investigations only to RCTs. They also do not report systematically on methodological issues such as blinding. Thus, the research in this section is often based on a variety of study designs, many of which could be considered of poor methodological strength.

Lane & Rollnick[80] aimed to assess whether the practice and rehearsal of communication skills leads to better outcomes following simulated patient and/or role-play training. Of the 23 studies reviewed 15 included a training intervention including simulated patients while the rest of the studies either included role-play training or communication skills training that was not further specified. They also provided limited comments on the rigour of the individual

studies assessed. As role-play was not included in the scope of the present review, only the information pertaining to simulated patients was extracted.

Ravert[81] did not provide detailed information on the rigour of the nine studies included in her analysis on the effectiveness of computer-based simulators. She however states a lack of strong conclusive studies and cautions the readers that the included studies did not document the reliability and validity of their evaluation instruments. It was noted that eight of the nine studies included in her review have a publication date of 1990 or earlier and only one study is from 2000. This raises a question about her search strategy and whether she has described adequately some of her inclusion criteria (e.g. it would have been useful to specify which modes of simulation qualify as 'computer-based' in her definition). From her comments it appears that she refers to human patient simulators, which combine mannequins with computer software, or to other interactive mannequins.

Two Supplementary reviews met the minimum critical appraisal criteria and are also shown in Table 3.1. These are:

- Gaffan, J., Dacre, J., Jones, A. Educating undergraduate medical students about oncology: a literature review. Journal of Clinical Oncology 2006; 24(12): 1932-1939
- Hamilton R. Nurses' knowledge and skill retention following cardiopulmonary resuscitation training: a review of the literature. Journal of advanced nursing 2005;51(3):288-297

Gaffan et al's[78] main aim was to review the literature on oncology teaching to undergraduate medical students i.e. they review a number of training modalities including models and standardised patients in medical students, while excluding dentistry and nursing students and including student satisfaction studies. Study designs are not clearly described and information on the rigour of individual studies is not always provided. The reviewers report

that the quality of the studies presented was variable with many of the studies being descriptive[78].

Hamilton's[79] main aim was to review the literature on the factors that enhance knowledge and skills following a number of different modalities of resuscitation training. She therefore reviewed a number of different training modalities including three studies on interactive mannequins and two studies using computer cardiac arrest simulation. However, the included studies are not clearly described in terms of study design, outcome measures used and limitations.

3.3.2.1. MODELS OR MANNEQUINS

The Core reviews by Lynagh et al[83] and Byrne et al[82] indentified two studies on mannequin intubation training with *in vitro* and *in vivo* results. The Supplementary review of Gaffan et al[78] identified three *in vitro* studies on teaching breast examination skills for cancer detection using models. The latter reviewers note that the intervention groups had previously used the models on which they were being tested, which left the control group at a systematic disadvantage. The Supplementary review by Hamilton[79] identified a further *in vitro* quasi-experimental study on CPR mannequin training. The results of the studies reviewed and the outcome settings are presented in Table 3.18 (Appendix 3.3).

Model training vs no training

In two clinical trials, comparing breast model training to no training led to significant improvements in lump detection for the intervention group (Table 3.18). The second study was an RCT but it was not made entirely clear what the intervention group teaching involved.

A further quasi-experimental study found that comparing self- instruction mannequin re-training to no such re-training led to positive results in terms of CPR performance on the mannequin.

Model training vs standard training

One RCT compared mannequin training to standard intubation training by tutor demonstration. Mannequin training led to better knowledge test scores *in vitro* but the researchers found no differences in self-rated intubation success rate over a 3 week clinical placement. A second prospective study compared a group that received mannequin training to a group of more experienced staff (standard training) and found no effects of training in terms of intubation success, times and complications measured in the field over 27 months (Table 3.18).

Comparing different types of model training

Two studies were found in the reviews which compared different types of model training *in vitro* (Table 3.18). The first study compared standard model teaching to dynamic model teaching of breast examination. Dynamic model teaching led to significantly higher lump detection in breast models. However, the reviewers did not describe what dynamic models teaching involved. In the second study, mannequin training only was compared to mannequin training plus independent mannequin practice plus feedback and the latter form of training was found superior in terms of intubation *in vitro*.

Byrne et al[82] reviewed three prospective studies that compared different types of model intubation training *in vivo* (Table 3.18). The first study compared an airway programme, which included mannequin practice, to the same airway programme also including a laryngoscopy video and found that the addition of the video led to better first attempt and overall intubation success rates *in vivo*. The second study randomised participants to mannequin training only or mannequin plus animal-based training or mannequin plus animal-based plus operation room training and found no effects of training in terms of intubation success, times and complications measured in the field over 27 months. The third study compared mannequin training to mannequin plus cadaver training and found no significant difference in intubation success rates in patients.

3.3.2.2. COMPUTER-BASED SIMULATORS

For the purposes of the present review the term 'computer-based simulation' in this section refers to interactive mannequins, computer software and patient simulators. The Core reviews by Ravert[81], Lynagh et al[83] and Byrne et al[82] identified seven RCTs, one cohort study, one prospective study with retrospective controls and seven studies with a pre-test post-test design pertaining to computer-based medical simulation. The supplementary review by Hamilton[79] reported on three additional studies, including one RCT, one cohort study and one study of unspecified design. The reviewed studies used different technical, cognitive and interactive skills as their outcome measures. Outcome measures included practical skills performance and knowledge. However, outcome measures are not always described and so it is not always clear whether these assessments were objective or subjective. The reviewed studies pertained to cardiology, anaesthesiology and emergency medicine. The results of the studies reviewed and the outcome settings are presented in Tables 3.19-3.22 (Appendix 3.3).

Computer-based simulation training using a pre-test post-test design

Of the nine studies covered in the included reviews, seven studies looked at the efficacy of computer-based training *in vitro* and two did not specify their outcome setting (Table 3.19). Five of the studies investigated cardiologic skills and knowledge, one study investigated cardiovascular arrest knowledge, two studies investigated CPR skills and one study investigated anaesthetic knowledge. All studies reported improvement in their outcomes following computer-based training with the exception of one-study which found no benefit in terms of CPR performance at 6 months following training.

Computer-based simulation training vs no training

One RCT found that comparing simulation training to no such training led to positive results in terms of trauma management *in vitro* (Table 3.20). No *in vivo* studies were found in the included reviews.

Computer-based simulation training vs standard training

Comparing computer-based simulation to other forms of training has led to inconsistent results *in vitro* in five RCTs, one prospective study with retrospective controls and one cohort study (Table 3.21). Three of the RCTs reported no differences in performance compared to video training (2 studies) and seminar-based teaching. The prospective study found that CD-ROM plus lab-based training led to inferior intubation performance compared to traditional training. Conversely, computer-based training led to superior performance on the simulator in two of the RCTs that compared this training to standard clinical experience and textbook study. The cohort study also favoured computer-based simulation in terms of knowledge and

skills *in vitro*, but this study compared cardiology electives with and without simulation training i.e. both groups had received the standard training.

Comparing computer-based simulation to other forms of training has led to positive results *in vivo* in one study (Table 3.21). The previously mentioned cohort study favored computer-based simulation used in addition to standard training in cardiology skills assessed in patients.

Comparing different types of computer-based training

One cohort study was cited that found 50 minutes of voice-activation mannequin (VAM) training led to better retention of CPR skills at 6 months than 20 minutes of VAM training (Table 3.22).

Integrative review of computer-based simulation training

Ravert[81] also performed an analysis of reported mean scores of education-related outcome measures (written examinations, psychomotor skills evaluations) by calculating effect sizes for each outcome in each study reviewed. Her results were based on seven pre-test post-test studies, one cohort study and one RCT. Based on her analysis 12 (75%) of the reported outcome measures favoured simulation, two (12.5%) did not support simulation and two (12.5%) were neutral. She concluded that computer-based simulation holds great potential for medical education but that more research is needed to further document its effectiveness.

3.3.2.3. STANDARDISED PATIENTS (SPs)

The Core paper by Lane & Rollnick[80] reviewed six RCTs, one clinical trial and eight studies of unidentified design pertaining to SP training for teaching communication skills. However, the setting in which outcomes have been evaluated is not always specified in this review. The Supplementary review of Gaffan et al[78] reported on five additional studies, including two RCTs, one clinical trial, one cohort study and one descriptive study on teaching oncology skills using SPs. However, additional teaching modalities were used in conjunction to SPs in a number of these studies. The results of the studies reviewed and, where available, their outcome settings are presented in Tables 3.23-3.25 (Appendix 3.3).

Simulated patient training vs no training

Research comparing SP training to no training *in vitro* generally led to positive results in the three studies included in this review. These studies assessed communication skills (3 studies; 2 favouring SPs) and examination skills (1 study) using simulated patients (Table 3.24). The sole exception was in a dental consultation study where SP training did not lead to significantly better communication skills. However, according to Lane & Rollnick[80] two of these studies did not include a baseline assessment.

Research comparing SP training to no training has also led to positive results in three RCTs *in vivo* (Table 3.24). Two of these studies investigated smoking-cessation counselling skills, including brief motivational interviewing techniques and active involvement of patients. The remaining study investigated a course that taught communication skills for oncologists using SPs in addition to other training methods.

Finally, four additional studies where the outcome setting was not specified led to positive results overall (Table 3.24). However, for three of these studies the reviewers reported they did not include a baseline assessment. Positive outcomes related to HIV counselling behaviours (2 studies), interviewing skills (1 study) and domestic violence consultation scores (1 study). The sole exception was in the domestic violence study where SP training did not lead to significantly better interpersonal skills.

Simulated patient training vs other forms of training

In terms of comparing SP training to other forms of training *in vitro* the research reviewed has led to inconsistent results (Table 3.25). Of the three clinical trials that compared the use of SPs to normal teaching for breast examination skills, two favoured the SPs in OSCEs. The third study, which was an RCT, led to inconsistent results as the SP trained group achieved higher sensitivity but lower specificity in detecting lumps in breast models. The reviewer did not specify what constituted normal teaching in this research. Furthermore, SP training in addition to video and lecture training also led to better lump detection in breast models compared to video and lecture alone. However, simulated patient training for smoking cessation did not lead to better consultation skills than role-play training in one RCT that used an SP assessment. Furthermore, in another RCT SP training did not lead to better communication skills than didactic lectures in speech pathology OSCEs. However, this last trial lacked a baseline assessment.

In contrast, two further studies that compared SP training to didactic lectures in unidentified outcome settings led to communication skills results that favoured SPs. Skills taught included breast examination and HIV risk assessment but the latter of the two studies did not include a

baseline assessment. HIV risk assessment and counselling skills were also the teaching objective of another study in which outcome settings were not specified by the reviewers. This was an RCT that found that SP teaching in addition to mailed educational materials led to better GP counselling and risk assessment practices than mailed educational materials alone. However, this study also lacked a baseline assessment.

In terms of comparing SP training to other forms of training *in vivo*, one RCT investigated a course which taught communication skills for oncologists using SPs in addition to other training methods. According to the reviewers, course attendance led to better communication skills with patients compared to providing written feedback alone.

Comparing different types of simulated patient training

One study was cited that compared pelvic examination teaching by a simulated patient to teaching by a physician using a simulated patient (Table 3.25). The reviewers reported that the SP trained group demonstrated similar technical skills and better interpersonal skills when assessed by an SP but also noted several methodological weaknesses including absence of baseline assessment and blinding.

3.3.2.4. OVERALL FINDINGS REGARDING SIMULATION TRAINING IN MEDICINE

No systematic reviews were found to specifically address simulation training for non-surgical skills. Thus, with the exception of Lynagh et al[83] the evidence presented for non-surgical skills was based on reviews of a less strict design that reported on research of variable study designs and quality. The overall conclusions resulting from the synthesis of the review evidence for non-surgical skills can be seen in Table 3.26.

Table 3.26. Overall summary of the review evidence on simulation vs other comparators pertaining to medicine

Intervention	**Author and Date**	**Comparator**				
		No training	Standard Training	Model training/ Physical trainer	Video-box training	Cadaver training
Model training	Gaffan et al 2006[78], Byrne et al 2008[82]	*In vitro*√	*In vitro*√ *In vivo* -	N.E.	N.E.	*In vivo* -
Computer-based training	Lynagh et al 2007[83], Hamilton 2005[79], Byrne et al 2008[82]	*In vitro*√	*In vitro*? *In vivo*√	N.E.	N.E.	N.E.
Simulated patients	Gaffan et al 2006[78], Lane & Rollnick 2007[80]	*In vitro*√ *In vivo*√ Unspecified setting √	*In vitro* ? *In vivo*√ Unspecified setting √	N.E.	N.E.	N.E.

(√ = Intervention superior, X = Intervention worse, ?= inconclusive evidence, - = no difference, N.E.=no systematically reviewed evidence found)

Models/ mannequins

- Model training vs no training (3 studies): Favourable results *in vitro*

However, Gaffan et al[78] reported that the training efficacy in their two studies was assessed in the same models. *In vivo* RCTs would be of benefit to confirm the effectiveness of models in teaching breast examination and other examination skills. SPs and hybrid simulation training (i.e. SPs attached to models) may represent useful choices in terms of comparators and assessment mediums.

- Mannequin training vs standard training (2 studies): Favourable results *in vitro*, no statistically significant differences *in vivo*
- Dynamic model training vs standard model training (1 study): Favourable results *in vitro* for the dynamic model group
- Mannequin training only vs mannequin training plus feedback plus independent practice (1 study): Favourable results *in vitro* for the mannequin plus feedback plus independent practice group
- Programme including mannequin training vs programme including mannequin training plus video (1study): Favourable results *in vivo* for the mannequin plus video group
- Mannequin training vs mannequin plus animal-based training vs mannequin plus animal-based plus OR training (1 study): No statistically significant differences *in vivo*
- Mannequin training vs mannequin plus cadaver training (1 study): No statistically significant differences *in vivo*

The studies reviewed did not provide clear evidence with regards to the effectiveness of mannequin training for clinical skills. Additional research may be of benefit to evaluate the effectiveness of mannequin training *in vivo* compared to other forms of training, including the respective effectiveness of various types of models, types of simulation, hybrid simulations and the additional technical features that lead to better training outcomes.

Computer-based simulators

- Computer-based simulation training without a comparator (9 studies): Favourable results *in vitro*
- Computer-based simulation training vs no training (1 study): Favourable results *in vitro*
- Computer-based simulation training vs standard training (7 studies): Inconclusive results *in vitro*
- Computer-based simulation training vs standard training (1 study): Favourable results *in vivo*

Further *in vivo* research is needed to determine the effectiveness of computer-based simulation in medical training, including the most effective 'doses' of training and resulting retention of skills.

Standardised patients

- SP training vs no training (9 studies): Favourable results *in vitro* and *in vivo*

However, the research reviewed was reported to have several methodological limitations. Further research of higher quality may be warranted to confirm the effectiveness of SPs in medical training.

- SP training vs standard training (10 studies): Inconsistent results *in vitro*, favourable results *in vivo* (1 study) and in unspecified outcome settings

Again, further high quality research, preferably *in vivo*, may be warranted to confirm the effectiveness of SPs versus other methods of standard training in relation to communication and examination skills.

3.3.3. FEATURES OF HIGH-FIDELITY SIMULATIONS THAT LEAD TO MOST EFFECTIVE MEDICAL AND SURGICAL TRAINING

In a Supplementary paper, Issenberg et al[16] reviewed 109 articles of various study designs searching for the features and uses of high-fidelity simulations that lead to most effective learning. Effectiveness was classified according to an expansion of the four Kirkpatrick training criteria (1. participation in educational experiences, 2a. change of attitudes, 2b. change of knowledge and/or skills, 3. behavioural change, 4a. change in professional practice, 4b. benefits to patients). Effective learning was defined as documented improvement in any of nine educational outcomes (clinical skills, practical procedures, patient investigation, patient management, health promotion, communication, information skills, integrating basic sciences, attitudes and decision making). High fidelity simulator categories included were realistic three-dimensional procedural simulators, interactive simulators and virtual reality simulators. The reviewers report that they 'did not evaluate whether simulators are more effective than traditional or alternative methods' but they selected articles that demonstrated 'in most cases, an improvement of knowledge, skills and attitudes'. Thus their review was classed as Supplementary.

Issenberg et al[16] found that the features and uses of high-fidelity medical simulations that led to most effective learning were: providing feedback on performance (51 studies), engaging in focused, repetitive practice (43 studies), curriculum integration (27 studies), engaging in a wide range of difficulty (15 studies), adaptability to multiple learning strategies (11 studies), capture of clinical variation (11 studies), controlled environment (10 studies), individualised learning (10 studies), defined outcomes (7 studies) and simulator validity in terms of fidelity (4 studies). They reported that the learning outcomes addressed in the studies reviewed were focused on practical procedure skills (over 75% of the studies) while learning outcomes in management skills, clinical skills and basic science knowledge were addressed in less than 20% of the studies. The reviewers stated that the direction of the evidence clearly shows that high-fidelity simulations facilitate learning when used under the right conditions and 'complement but do not duplicate education in real settings involving real patients'.

3.3.4. EFFECTS OF HOURS OF SIMULATION PRACTICE ON LEARNING OUTCOMES IN HIGH-FIDELITY MEDICAL EDUCATION

In a supplementary paper, McGaghie et al[84] used a subset of 32 quantitative experimental or quasi experimental studies from the Issenberg et al systematic review, where a high-fidelity medical simulator had been used. These were 20 randomised trials (14 used psychomotor outcomes, two used cognitive outcomes and four used both psychomotor and cognitive outcomes) and 12 cohort studies (five used psychomotor outcomes, two used cognitive outcomes and five used both psychomotor and cognitive outcomes). Included studies pertained to surgery (9), anaesthesiology (8), cardiology (2), emergency medicine (1), nursing (2), chiropractic specialty (1), paediatrics (2), dentistry (1), paramedics (1) and family

medicine (1) and hence there were a variety of outcome measures. Sample sizes ranged from 10 to 208 students.

Studies pertaining to surgery and medicine were both included in the review above. The literature reviewed included six of the studies by Haque & Srinivasan[71], 18 of the studies by Sutherland et al[77] and four of the nine studies by Ravert[81]. However, McGaghie et al[84] did not address simulation efficacy and effectiveness directly. Instead they focused on the effect of hours of simulation practice on learning outcomes. Thus their review was classed as Supplementary.

McGaghie et al[84] extracted data from each study to measure the intensity of the simulation intervention, which they coded in five categories in the form of hours of simulator practice. One of the categories included 11 studies where no such data had been reported, five studies were coded in a 0-1 hour category of simulation practice, nine studies coded in the 1-3 hours category, five studies coded in the 3.1-8 hours and two studies coded in an 8 plus hours category. A weighted effect size, accounting for the number of cases represented by each learning outcome variable, was calculated to standardise each outcome variable in the 32 studies and random effects models were used to cast the heterogeneous learning outcome measures into a common metric. Using ANOVA the authors found that 'repetitive practice involving medical simulations is associated with improved learner outcomes' and that 'more practice yields better results' in a 'dose-response' manner. However, they caution readers that the results of the analysis were 'strongly influenced by the two studies where practice time met or exceeded eight hours' although this does not suggest diminishing returns to training that might be expected.

3.4. DISCUSSION

3.4.1. SUMMARY OF THE FINDINGS

This review attempted to provide an overview of the current review evidence on the efficacy and effectiveness of simulation training of technical and non-technical skills in healthcare. Based on the review evidence identified, none of the included surgical and medical simulation training methods has yielded unequivocal results in favour of simulation training over other types of training. This has also been the conclusion of five of the six systematic reviews identified[16, 76, 77, 82, 83]. However, it has been suggested that the absence of significant differences in some cases may be an indication that the simulation training under investigation can be as good as other methods[71, 77]. Only five studies were found to have reported that simulation training led to a significantly worse outcome than the comparator. A seemingly paradoxical result was that the overall *in vivo* evidence comparing surgical model training to no training appeared inconclusive while the *in vivo* evidence comparing surgical model training to standard training indicated that surgical model training was superior to standard surgical training. This could be due to a number of factors such as study design limitations, different outcome measures being used, the quality of the training and the type of skill being taught.

Furthermore, meta-analysis has yielded some positive results. Based on *in vivo* studies, VR training appears to lessen the completion time of surgical tasks[71]. The MIST-VR simulator has been used in a large proportion of the reviewed VR studies. The MIST-VR is a system that teaches basic psychomotor skills to perform a laparoscopic operation, as well as skills required in advanced laparoscopic procedures[88]. It has been found to be reliable and valid in numerous studies and can be used for simple and abstract testing and acquisition of

psychomotor skills but does not test cognitive knowledge or complete operations[89]. Minimally invasive operations can be easier to simulate due to the limited visual and haptic feedback involved[90]. However, the evidence of the role of VR simulators in the transfer of skills to the operating room remains equivocal.

Nevertheless, based on the studies found in the included reviews VR training appears to be preferable to no simulation-based training. This has also been the conclusion of the corresponding reviews. However, what reviewers classify as 'no training' can be ambiguous at times. In many studies simulation-based training was in addition to normal training programs, thus 'no training' actually may mean no simulation-based training.

Upon observation of the type of procedures being assessed in the individual studies in the reviews it would appear that studies investigating simulation training on surgical laparoscopic procedures tended to yield positive results more often than studies investigating simulation training on procedures which could be perceived as less technically demanding, namely IV catheterisation and intubation. This prompted the hypothesis that the effectiveness of simulation training depends on the degree of difficulty of the task with simulation being more effective as the difficulty of the task increases.

In addition, with regards to the effect of hours of simulation practice on learning outcomes, preliminary analysis has shown that more high-fidelity simulation practice appears to lead to better results in a dose-response manner[84]. However, the same is likely to be true across all

media, not just simulation. Furthermore, the research to date precludes any conclusions with regards to whether simulation training leads to skills retention over time[83].

According to Issenberg et al[16], the direction of the evidence shows that high-fidelity simulations facilitate learning under the right conditions and complement education involving real patient contact. According to the same researchers the weight of the literature points towards providing feedback, repetitive practice, curriculum integration, range of difficulty, multiple learning strategies, capture of clinical variation, controlled environment, individualised learning, defined outcomes and simulator validity as the features and uses of high-fidelity medical simulations that lead to the most effective learning[16]. This may be an indication that the features of the simulation rather than the type of simulator determine training effectiveness.

3.4.2. LIMITATIONS OF THIS REVIEW

The present review has several limitations. The initial literature search of six databases revealed a dearth of systematic reviews, while the majority of the reviews found following the extended search in MEDLINE and EMBASE were overviews that did not seem to use a systematic approach. This precluded a more systematic presentation of the evidence provided including quantitative synthesis of results. In addition, the final search was confined to two databases and the search terms used may not have been adequately specific to cover all the available reviews on simulation healthcare training.

A considerable portion of the information found was part of reviews covering many different methods of training in one medical field rather than simulation methods alone or covering a simulation topic that did not match the exact topic of this review of reviews. The critical appraisal methodology used led to the exclusion of the majority of these reviews as it was felt that their inclusion would not lead to a balanced view of the evidence. In addition, some of the studies that were included in this review would not pass a strict critical appraisal, although these were deemed useful in adding further insights regarding the efficacy or effectiveness of simulation training.

3.4.3. LIMITATIONS OF THE REVIEWS INCLUDED AND OF THEIR COMPONENT STUDIES

Most of the reviews encountered in the literature search targeted the fields of surgery and anaesthesia and seemed to consist of general overviews, often descriptive in nature and with no search strategy reported. In agreement with Aucar et al[74], systematic citation of original data to support validity was notably absent and the majority of the reviews encountered were feasibility, editorial and theoretical comments. The majority of review reports on medical simulation effectiveness do not have measurable outcomes explicitly stated, but rather group together theory, opinion, simulation descriptions, history and trials of various study designs[16]. A more systematic approach by future reviewers might help in acquiring a clearer picture of existing research on the training effectiveness of the various types of simulation used by different medical specialties and to indentify respective research gaps.

In addition there seems to be a lack of clarity on the terminology used to describe the different types of simulators. For example, Ravert[81], Gerson[21] and Sutherland et al[77] all refer to

'computer-based simulation' and 'computer simulation'; however, the first reviewer has taken their term to mean computer software and patient simulators, while the other two reviewers have used this categorisation for VR simulators. This can be confusing for readers who do not have extended knowledge on the different types of computer-based simulation. Further confusion can result from the lack of a common approach regarding the classification of comparators and the interpretation of individual study results. For example, two high-quality systematic reviews, Sutherland et al[77] and Lynagh et al[83] appear to have had different definitions of what constitutes 'no training' and 'standard training' resulting in the same studies found being classed differently with regards to these comparators on two occasions. These same reviews were also found to disagree on the interpretation of the same studies' results on three occasions with one review reporting 'no difference' where the other review had reported superiority of one comparator over the other.

Furthermore, simulator validation studies appear to be focused on laparoscopic simulators and few of the remaining studies specifically address the validity and reliability of their assessments[64, 74]. Lack of standardisation has been noted in operative techniques used and a need for development of a core set of objective measures of operative skill has been expressed in order to confirm the role of simulators in laparoscopic surgery[77, 91]. One of the underlying reasons may relate to McGaghie et al's[84] observation that 'most published work demonstrates a lack of awareness about basic designs used for research in education, behavioural science and the clinical disciplines'. Arnold & Farrell[73] suggest that a more interdisciplinary approach involving also psychologists, ergonomists, computer scientists and other disciplines may be a way forward.

Investigating the impact of simulation on performance requires an understanding of the concepts of validity, reliability and skills-transfer testing[92]. Limited data supporting the construct validity of simulator surgical training have been reported suggesting that VR and bench simulators *may* have value for surgical training[74]. However, according to Mantovani et al[93] most of the VR applications can be considered 'one-off' creations tied to a proprietary hardware and software, which are difficult to use in contexts other than those in which they were developed. As Khalifa et al[94] has noted, the findings of individual trials speak primarily for the specific VR trainer tested and not for VR trainers in general. This lack of generalisability may also be the case for the other methods of simulation training reviewed.

Furthermore, as Jha et al[36] have noted, studies of the effectiveness of simulators are often limited in that they measure performance using the same training simulator and this may favour those who have trained on the simulator and may not translate to actual patient care. Proficiency in the simulator does not ensure proficiency in clinical settings[74]. However, examining how proficiency in the simulator correlates with proficiency in reality remains an underexplored field. It has been noted that the way to measure learning transferability in medicine has not been studied and the transfer efficiency rate of how much time spent in the simulator is equivalent to time learning on the real task remains unknown[1, 95].

In addition, the quality of the primary research that has been used to compose the review evidence found in this review limits the quality of the findings presented, a point previously noted by Issenberg et al[16]. The range of different study designs that have been employed with their respective strengths and limitations, the lack of blinding in a considerable proportion of

comparative studies and the generally small sample sizes further compound the evidence reported.

3.4.4. CONCLUSIONS

Further high quality and sufficiently powered *in vivo* research is needed to investigate the effectiveness of surgical and medical simulators including inanimate and ex-vivo models, video-box trainers, VR simulators, cadavers, animals, computer-based patient simulators and standardised patients. Further *in vivo* comparisons of different simulators may be useful in order to determine the optimum form for different skills and to avoid using high-cost forms in cases where a simpler model or role-play may be as effective. Further research may also be warranted in order to determine the optimum format of training in individual simulators in terms of level of difficulty, type of feedback, frequency and duration of practice sessions and maintenance of skills in the longer-term.

Furthermore, as Jha et al[36] and Haque & Srinivasan[71] have previously pointed out, among the studies reviewed there was a dearth of studies evaluating the impact of simulation training on actual patient care, medical error or showing a clear link between simulator training and patient outcomes. In addition, the majority of reviews in this report addressed simulation training studies that evaluated the acquisition of psychomotor skills. No systematic reviews were found on the acquisition of cognitive skills such as decision making and interactive skills in terms of working in a team. The gaps in these topics may be an indication that a systematic review and or/further research is also needed in these domains. Future reviewers should seek to adopt a common approach with regards to the simulation typology and

classification of comparators used so as to avoid confusion of the reader with regards to the type and direction of the existing evidence.

CHAPTER 4

THE EFFECTS OF LARYNGEAL MASK AIRWAY PASSAGE SIMULATION TRAINING ON THE ACQUISITION OF UNDERGRADUATE CLINICAL SKILLS: A RANDOMISED CONTROLLED TRIAL

4.1. INTRODUCTION

The laryngeal mask airway (LMA) can be used in a variety of airway management situations including the administration of inhaled anaesthetics, as an airway for controlled or spontaneous ventilation during the administration of anaesthesia, as a tracheal intubation assistive device and as an emergency airway device during resuscitation[96]. In many circumstances it is an alternative to the more technically demanding process of intubation. Effective use of the LMA requires learning proper insertion technique in normal patients undergoing routine surgical procedures with general anaesthesia[96]. However, there is a move towards simulator training for learning practical clinical skills[2, 97, 98] and hence Dierdorf recommends practice on a mannequin before attempting the technique of LMA insertion on real patients[96]. This chapter reports on an RCT which investigated the effectiveness of (and hence necessity for) such training. The ideal study would:

1) Evaluate effectiveness on real patients rather than on the mannequins themselves (at least until it could be shown that *in vitro* testing on a mannequin is a reliable surrogate for *in vivo* proficiency).

2) Compare outcomes in an intervention group with outcomes in a randomly generated control group.

In the process of systematically reviewing simulation training four studies were found that had measured the outcomes of LMA training *in vitro* (i.e. using a mannequin) but with no contemporaneous controls[99-102]. These studies fulfilled neither of the above criteria (*in vivo* testing of proficiency and concurrent controls). Three studies had measured outcomes on real patients but these did not use contemporaneous controls[103-105] thereby fulfilling only the first of the above criteria. Only one small study evaluated mannequin training *in vitro* using contemporaneous controls[106].

4.1.1. UNCONTROLLED STUDIES *IN VITRO*

In a cohort study, Ander et al[99] assessed resident rescue training (n=40) using three airway devices, including the LMA. Training consisted of lectures, reviewing scenarios, videos and hands-on demonstrations on a mannequin. Their outcome measure was 'time to ventilate an airway mannequin'. Time to successful ventilation using the LMA was reported at 6.9 sec immediately after training with a modest increase in mean time at 6 and 12 months. The authors also reported that previous and interval experience did not affect performance. However, this study does not seem to have included a baseline assessment. Furthermore, performance was assessed on bench models rather than actual patients.

In a cohort study, Tiah et al[100] compared the skills retention of three airway devices, including the LMA, in 93 medical students with a view to recommend changes to the medical school curriculum. Training consisted of mannequin training in small groups. Their outcome measure was time to successful insertion, number of attempts required and complications encountered. Time to successful insertion using the LMA was reported at 25.4 sec immediately after training with a 13.5 sec increase in mean time at 6 months ($P<0.001$). First-attempt success rate was 96% at 0 months and 92% at 6 months ($P=0.549$). Overall complication rate was 3% at 0 months and 10% at 6 months ($P=0.688$). However, this study does not seem to have included a baseline assessment. Furthermore, performance was assessed on mannequins rather than actual patients.

In a prospective study, Vertongen et al[101] compared the skills retention of the LMA and the oesophageal-tracheal Combitube over six months. One hundred and one nursing, medical and theatre staff were taught LMA insertion on an Ambuman mannequin. Training consisted of a written handout, a video, a demonstration and mannequin practice. Time to successful ventilation was recorded between one to 19 days following training and at a minimum of six months for 86 participants (85%). At the initial testing 90% were successful at LMA ventilation while 85% were successful at retesting, a non statistically significant decline. However, the authors noted that the skill retention for the LMA may actually reflect the inherent simplicity of the device. Median time to insertion was 28sec on initial testing and 24 sec on retesting.

In a prospective study, Weksler et al[102] compared the skills retention of the LMA, the endotrachelal tube and the Combitube over six months in 5th year medical students. Training consisted of an audio-visual presentation of theory and proper insertion techniques and a demonstration of the techniques on a mannequin. First attempt success rate was recorded following training, and at six months, for 47 participants. At the initial testing 100% were successful at LMA ventilation while 93% (n=29) were successful at retesting. The authors concluded that using mannequin training LMA insertion skills are easily learned and better retained than those for endotracheal intubation.

4.1.2. UNCONTROLLED STUDIES *IN VIVO*

Davies et al[103] trained 11 naval medical trainees using instruction through videotapes, mannequin practice, and a demonstration on an anaesthetised patient. Subsequently all participants undertook LMA placement in 10 patients scheduled for routine elective surgery. The choice of patient where trainees were tested on was controlled i.e. ASA 1 (American Society of Anaesthesiologists physical status) patients with no loose teeth or crowns. The researchers reported 100% success in the participants' first patient in terms of LMA insertion, 82% success for the second patient, above 90% subsequently and an overall success rate of 94%. The mean insertion time was 20 sec. The authors reported a lack of a significant learning curve.

Frascone et al[105] trained 11 helicopter flight paramedics and nurses on LMA placement using a didactic and mannequin-based training session. Participants then placed two LMAs on consecutive adult patients who were undergoing surgery with an interdental space of less than 3 cm. The researchers reported 100% LMA placement success rate in all 22 patients and

found no learning effect when comparing mean time to ventilation for the first and second LMA placement (p=0.45). Average time to ventilation across all LMA placements was 36.8 seconds with second placement mean time being faster than the first time.

Murray et al[104] trained 208 paramedics in the use of LMA using a 2-hour self-preparation module and training video followed by 4 hours of didactic teaching and classroom practice on an ALS Trainer Laerdal mannequin. They found a high classroom success rate (100%) in correct mannequin LMA placement. However, this translated to a moderate success rate (64%) in LMA insertion and ventilation in 283 prehospital adult non-traumatic cardiac arrest patients over a period of 13 months. Success was determined using subjective evaluations by the paramedic at the scene and by the emergency physician or respiratory therapist on arrival in the emergency department. According to the investigators the frequency of success did not vary significantly according to experience with insertion of the LMA.

4.1.3. CONTROLLED STUDIES *IN VITRO*

Morse et al[106] compared three types of LMA placement training (demonstration alone, demonstration plus limited (5 attempts) practice on a mannequin or demonstration plus extended (10 attempts) practice on the mannequin) among 35 dental students. The outcome was measured neither on a mannequin, nor on 'real' patients, but on a cadaver. A fiberscope was used to grade the quality of the LMA placement. Students who practiced on a mannequin achieved significantly shorter LMA insertion times and superior placement grades than those who received only a demonstration. Placement grades for the group that had practiced 10 times on the mannequin were not statistically significantly different from the group that had practiced 5 times. It was unclear whether this was a randomised study.

Thus no studies were found that fulfilled both the criteria of use of randomised controls and measurement of proficiency on real patients. The current chapter describes such a study. Currently, medical students at the University of Birmingham in the UK receive only very limited instruction on LMA placement and the majority have one or two practice attempts on a mannequin. This study set out to test the hypothesis that an additional session of formal simulation training would promote speed of learning and result in a higher level of skill than brief simulation training when LMA placement is first undertaken in clinical practice. An additional objective was to compare self-assessment of success in this particular procedure with objective assessment by a third party in a clinical setting.

4.2. METHODS

Participants

The study was approved by the West Midlands Multi-Centre Research Ethics Committee and was undertaken during the period August 2006 to March 2007. Eligible participants were all Year 4 University of Birmingham medical students who were about to undergo their clinical attachment in Anaesthesia, Respiratory and Intensive Care Medicine (ARICM) for the academic year 2006-2007. One hundred and twenty-six students across 5 blocks volunteered to take part after listening to a talk describing the study and each participant gave written consent. Participants completed a brief baseline questionnaire recording their demographic features and any previous LMA placement training that they had received (Appendix 4.1). All participants had received the standard brief mannequin training in LMA placement that is a compulsory part of their ARICM module. This typically involves a demonstration of the technique by a clinical instructor on a mannequin followed by one or two attempts by the trainees.

Baseline assessment

Each participant undertook a baseline assessment at The University of Birmingham Medical School. This consisted of performing LMA placement on a mannequin (Laerdal Adult Airway Trainer) once using a size-4 LMA Classic™ (Fig. 4.1). Two anaesthetists (a different pair in each block) rated each student's performance on a pro-forma (Appendix 4.2.). The time to ventilation success or failure was recorded. Ventilation success was verified by direct visualisation of chest expansion of the mannequin with bag-tube-ventilation. In the cases where the two raters did not agree on ventilation success, the attempt was recorded as unsuccessful. The anaesthetists rated the participants' handling of the LMA and their overall success in LMA placement using a 5-point Likert-type scale. Twenty-two of the participants of the first block consented to video-recording of their baseline attempt. These recordings were used to investigate the test-retest agreement of the anaesthetists. A flow chart of the assessments used in this RCT is presented in Fig. 4.2.

Figure 4.1. Participant undertaking a baseline assessment

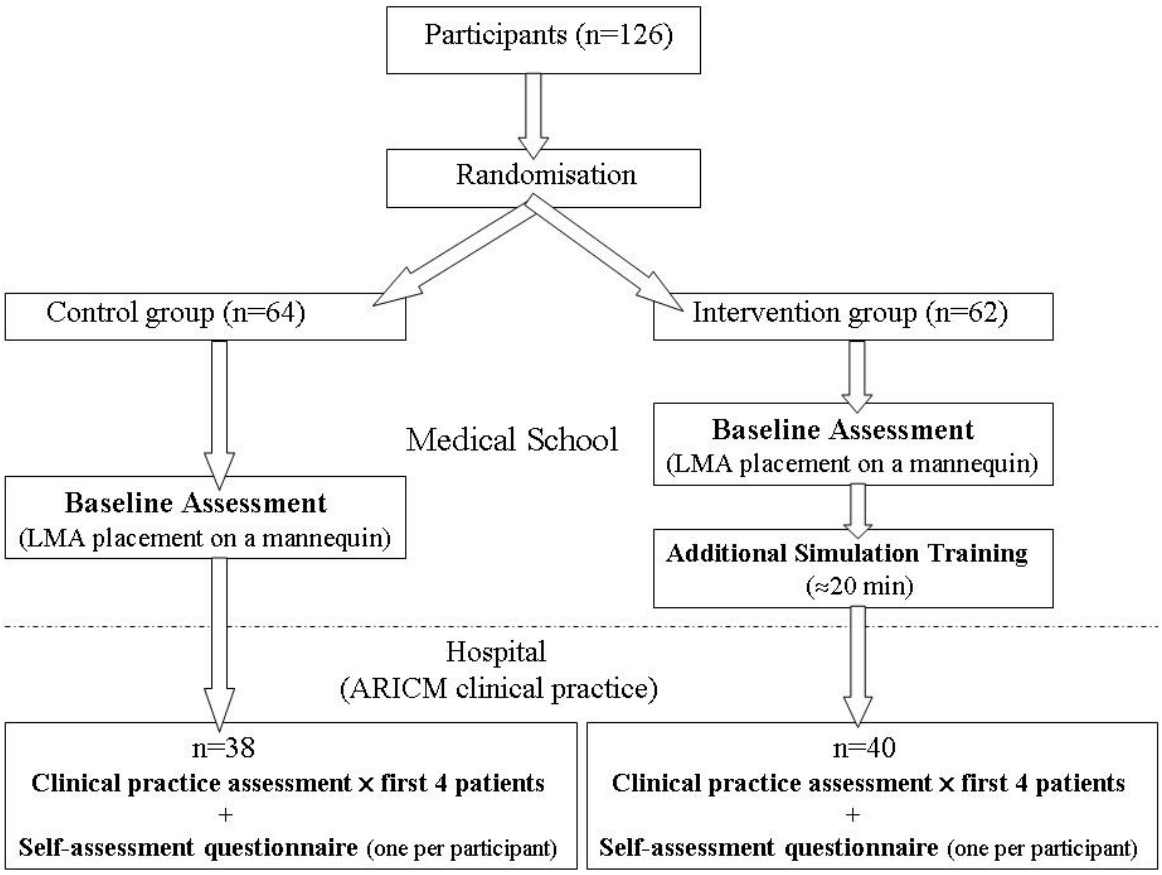

Figure 4.2. Flow chart of the RCT's assessment process

Method of randomisation

Following the baseline assessment, each participant was asked to open an envelope containing their group allocation. Block randomisation and sequentially numbered sealed opaque envelopes were used for the allocation. A random-number table and blocks of 4 were used to generate the random allocation sequence[107]. The researcher administering the allocation (Celia A Brown) was independent of the recruitment process (by Elpiniki Laiou) and did not have any knowledge of the participants' baseline characteristics or assessment. All participants were given a pack containing 4 sequentially numbered assessment forms (Appendix 4.3.) in sealed envelopes, a clinical practice self-assessment questionnaire (Appendix 4.4.) and written instructions on how to fill these in (Appendix 4.5.).

Intervention

Participants in the control group received no additional mannequin training. Participants in the intervention group received approximately 20 minutes of additional LMA placement training on Laerdal Airway Management Trainers, administered at the end of the baseline assessments. The participants were taught the use of the LMA (size 4, LMA Classic™) on the mannequin in groups of 4-8 each. The training consisted of a step-by-step demonstration of the LMA placement technique on the mannequin by an anaesthetist and supervised practice of the technique until the participants had demonstrated a correct mannequin LMA placement. Training was in accordance with the instruction manual [108]. Participants were encouraged to practice for as long as they needed.

Clinical practice assessment

All participants subsequently spent six weeks undertaking their standard ARICM clinical training in one of 11 hospitals (Table 4.1). The anaesthetists supervising the participants' LMA placements following induction to anaesthesia in the operating room were asked to fill in an assessment form (Appendix 4.3) on the participants' first 4 attempts to insert the LMA in patients. The form contained questions about placement success and the following data were collected: 1) rating of the overall success (the primary outcome) of the LMA placement on a 5-point Likert-type scale (1 being 'extremely poor' and 5 being 'excellent') 2) successful ventilation following the LMA placement, 3) rating of the handling of the LMA during the insertion on a 5-point Likert-type scale, 4) whether time to successful placement was less than 40 seconds and 5) the number of insertion attempts. Successful ventilation was determined clinically by observation of adequate seal, satisfactory chest movement and observation of a normal capnographic curve where applicable. The anaesthetists conducting the assessments

were ‘blind’ to group assignment and the participants had been asked not to reveal their group allocation.

Table 4.1. List of the participating hospitals

Alexandra Hospital, Worcestershire Acute Hospitals NHS Trust
City Hospital, Sandwell and West Birmingham Hospitals NHS Trust
Good Hope Hospital, Good Hope Hospital NHS Trust
Heartlands Hospital, Heart of England NHS Foundation Trust
Manor Hospital, Walsall Hospitals NHS Trust
New Cross Hospital, Royal Wolverhampton Hospitals NHS Trust
Queen Elisabeth Hospital, University Hospital Birmingham NHS Foundation Trust
Russell's Hall Hospital, Dudley Group of Hospitals NHS Trust
Sandwell District General Hospital, Sandwell and West Birmingham Hospitals NHS Trust
Selly Oak Hospital, University Hospital Birmingham NHS Foundation Trust
Worcester Royal Hospital, Worcestershire Acute Hospitals NHS Trust

The participants were also asked to fill in a self-assessment questionnaire on their first 4 LMA placements in patients. The questionnaire asked participants to rate their own handling of the LMA and their overall success in LMA placement for each patient on 5-point Likert-type scales identical to those used by the anaesthetists. It also asked them to comment on the usefulness of their respective mannequin training.

Sample size

Based on consultation with the Year 4 Lead of the ARICM module, the probability of receiving an overall LMA placement success rating of >3 (i.e. above average) on the corresponding Likert-type scale, for a student's first patient, was set at 0.50 for the control group and at 0.75 for the intervention group. Based on 80% power to detect a statistically significant difference ($\alpha = 0.05$), 55 participants were required for each study group. To compensate for drop-out/failure to complete, the planned number of participants had been 85 participants per group.

Analysis

Statistical analysis was performed using SPSS 12.0.1 and R 2.6.2 for Windows. The distribution of data was determined using the Shapiro-Wilk test of normality. Pearson Chi-squared analyses were used to compare groups in terms of rates of achieving effective ventilation, establishing ventilation in ≤40sec and achieving effective ventilation at the first insertion attempt. Fisher's Exact Test was used when expected counts were less than 5. Fisher's randomization T-test was used to compare groups in terms of LMA handling and overall LMA placement success ratings of 1-5. A one-sided test was used as it was hypothesised that the additional simulation training would not result to poorer skill than brief simulation training. The Mann-Whitney U test was used to compare successful attempt times in the baseline assessment as data were not normally distributed. A p-value of <0.05 (one-sided) was considered statistically significant. Baseline inter-rater agreement, baseline intra-rater agreement and clinical practice instructor-participant agreement with regards to whether they awarded an overall LMA placement success rating of <3, 3 or >3 were explored using the kappa statistic[109]. The relationship between instructor and participant overall LMA

placement success ratings was analysed using the Spearman's ρ (rho) correlation coefficient. Pair wise deletion was used with regards to missing data. The participants' quotes on the self-assessment questionnaire were categorised into common themes.

4.3. RESULTS

4.3.1. POPULATION

One hundred and twenty six Year 4 medical students (34%) enrolled in the study and participated in the baseline assessment (Table 4.2).

Table 4.2. Participant characteristics at baseline and follow-up

	At baseline		At follow-up	
Study group	**Intervention**	**Control**	**Intervention**	**Control**
No of participants	62	64	40	38
Age mean (SD) years	23.2 (1.4)	23.1 (1.1)	23.5 (1.6)	23.0 (0.9)
Sex: Male	20 (32%)	23 (36%)	11 (28%)	12 (32%)
Female	42 (68%)	41 (64%)	29 (72%)	26 (68%)
Ethnicity:				
British White	44 (71%)	37 (58%)	31 (77%)	22 (58%)
British Asian	9 (14%)	11 (17%)	4 (10%)	6 (16%)
British other	2 (3%)	6 (9%)	0 (0%)	4 (10%)
Non British	6 (10%)	10 (16%)	4 (10%)	6 (16%)
Not disclosed	1 (2%)	0 (0%)	1 (3%)	0 (0%)
Prior LMA practice:				
Mannequin ARICM	43 (69%)	43 (67%)	28 (70%)	28 (74%)
Other mannequin	3 (5%)	4 (6%)	2 (5%)	3 (8%)
Patient	7 (11%)	12 (19%)	4 (10%)	8 (21%)
None	17 (27%)	15 (23%)	11 (28%)	6 (16%)

Overall, 78 (62%) of the participants completed the study by returning some or all of their instructors' assessment forms from their clinical practice in the hospitals (Fig. 4.3). However, due to missing data the response rate to each of the 4 individual assessments was variable. Three additional participants returned their clinical practice self-assessment questionnaires, giving a total of 81 participants who returned some form of data at the end of the study. Table

4.3 contains data on the participants' demographic aspects at follow-up versus those of the overall MBChB Year 4 population. Data on nationality were used as data on ethnicity could not be obtained for the overall population. The study population and the overall MBChB Year 4 population were similar in terms of mean age and nationality and their difference in terms of sex was small (Table 4.3).

Table 4.3. Participant characteristics at follow-up versus overall MBChB Year 4 population characteristics

Group		Study participants	Overall population
No of participants		78	393
Age mean (SD) years		23.2 (1.3)	23.4 (1.8)
Sex:	Male	23 (29%)	158 (40%)
	Female	55 (71%)	235 (60%)
Nationality:	British	68 (87%)	358 (91%)
	Non British	10 (13%)	35 (9%)

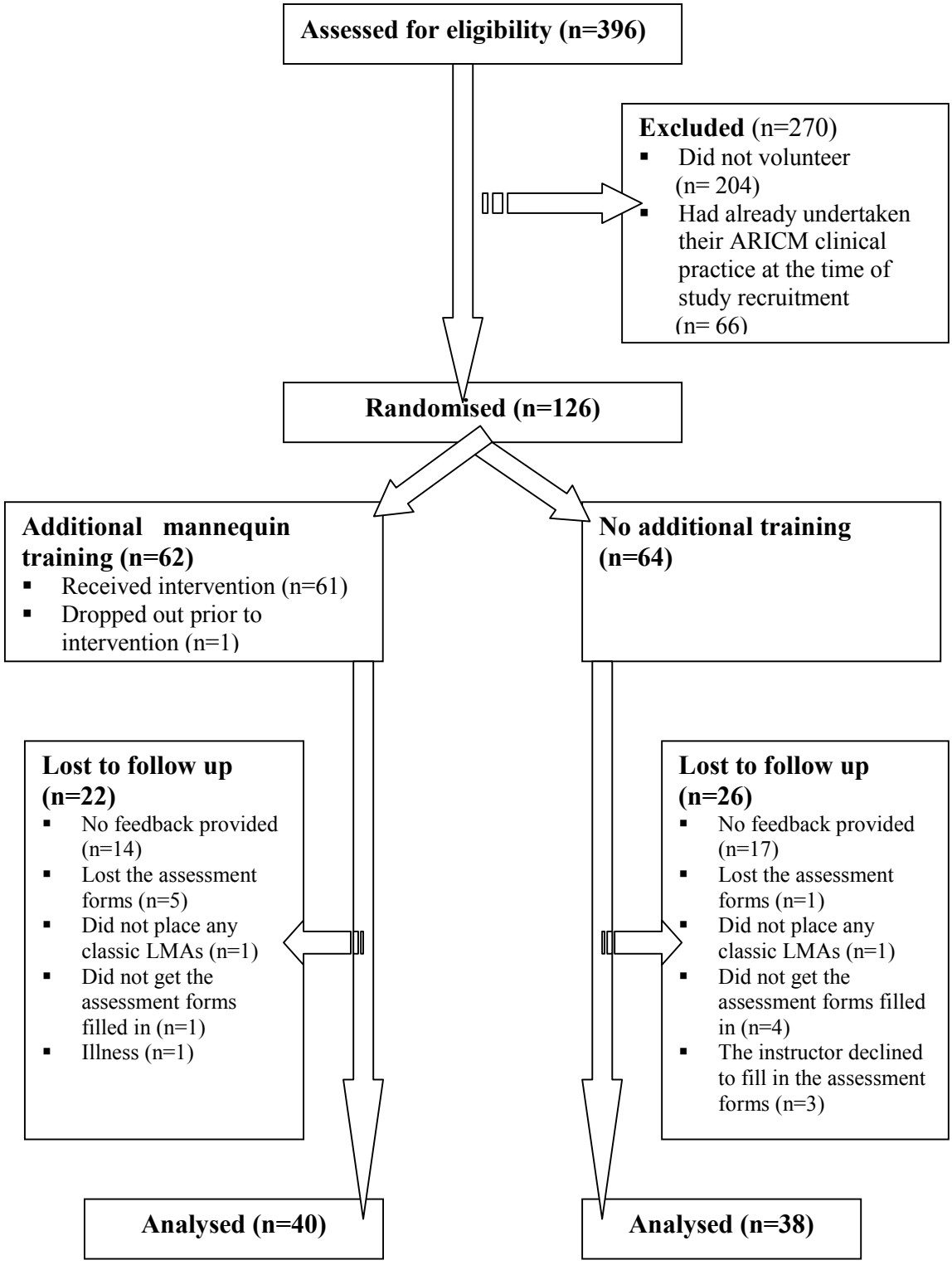

Figure 4.3. Participant flow chart following CONSORT scheme

4.3.2. PRIOR EXPERIENCE

One hundred and twenty-six participants (100%) filled in a baseline questionnaire on their previous experience in LMA placement. Eighty-six participants (68.2%) had practiced LMA placement on a mannequin as part of their medical school ARICM module. Eighteen of these participants reported that they had also had the opportunity to practice LMA placement on a Trauma course (n=4) or on a patient (n=14).

Eight participants (6.3%) reported that they had not practiced LMA placement on a mannequin as part of their ARICM module at the time of the baseline assessment but had had the opportunity to practice LMA placement on a patient (n=5) or on a mannequin (n=3).

Thirty-two participants (25.4%) reported that they had never practiced LMA placement at the time. However, 3 of these reported having had a verbal tutorial. In two of these cases the participants had had their ARICM mannequin training but had not practiced LMA placement on the mannequin.

The primary reason that participants had not had their ARICM mannequin practice at the time of the baseline assessment was that a small number of hospitals carried out their ARICM mannequin session in a later date than the one originally anticipated, while one participant reported having missed the session.

4.3.3. BASELINE ASSESSMENT

All 126 participants (100%) had one attempt in LMA placement on a mannequin. 66% achieved successful ventilation (Table 4.4). Ventilation success was 69% for participants who had had any LMA placement practice and 56% for participants who had not had previous LMA placement practice (X^2= 1.767, p=0.184). The mean ±SD time taken to achieve successful ventilation was 33.2 ±13.3 seconds in the group who had had LMA practice compared to 31.9 ± 10.1 seconds in the group who had not had previous LMA practice (Mann-Whitney U= 578.5, p=0.943).

Table 4.4. Baseline assessment success rate and successful ventilation times with and without previous LMA placement practice

	LMA practice	No LMA practice	Total
Participants (No.)	94	32	126
Ventilation Success (%)	69.1%	56.3%	65.9%
Time (mean± SD) sec	33.2±13.3	31.9±10.1	32.9±12.7

No statistically significant differences were found between groups at baseline in successful ventilation, LMA handling mean ratings, overall success mean ratings and time taken (Table 4.5).

Table 4.5. Baseline assessment success rates and successful ventilation times

	All participants (N=126)				Participants followed-up (N=78)			
	I*	C*	Test statistic	P-value	I	C	Test statistic	P-value
Overall success Mean rating	2.88	2.70	$\bar{I} - \bar{C}$ 0.18	0.172†	2.98	2.68	$\bar{I} - \bar{C}$ 0.30	0.122†
Overall success rating > 3	22.6%	18.8%	X^2 0.282	0.595	25.0%	18.4%	X^2 0.495	0.482
Successful ventilation	62.9%	68.8%	X^2 0.479	0.489	67.5%	65.8%	X^2 0.026	0.873
LMA handling mean rating	3.02	2.98	$\bar{I} - \bar{C}$ ** 0.04	0.453†	3.10	3.05	$\bar{I} - \bar{C}$ 0.05	0.450†
Mean time (SD) sec	33.6 (12.2)	32.3 (13.2)	Mann-Whit. U 787.5	0.519	33.9 (12.7)	29.1 (11.7)	Mann-Whit. U 262.0	0.165

* I = Intervention group, C= control group

** $\bar{I} - \bar{C}$ = difference between means

† P-value obtained using Fisher's randomisation t-test

The inter-rater agreement using the Likert-type overall success rating scale was fair with kappas ranging between 0.22 and 0.33 (Table 4.6). Inter-rater agreement in terms of above average (>3) overall success ratings was fair to moderate ranging between 0.28 and 0.47. With regards to ventilation success, the inter-rater agreement was moderate to very good with kappas ranging between 0.43 and 1.00. The inter-rater agreement using the Likert-type LMA handling rating scale was poor to fair with kappas ranging between 0.17 and 0.33.

Table 4.6. Anaesthetists' agreement (kappas) at baseline

ARICM block	Overall success rating 1-5	Overall success rating > 3	Successful ventilation (yes/no)	LMA handling rating 1-5
1 (n=32)†	0.22	0.45	0.50	0.17
2 (n=26)	0.32	0.29	0.66	0.24
3 (n=30)	0.33	0.28	0.70	0.33
4 (n=20)	-*	0.47	0.43	0.25
5 (n=17)	-*	-*	1.00	-*
Total (N=125)	0.25	0.32	0.61	-*

* Kappa statistics could not be computed due to asymmetry of the result 2-way tables.

† One missing value

The intra-rater agreement of the Likert-type ratings of the two raters of the 1st ARICM block at six months was fair for overall success, moderate in terms of above average (>3) overall success ratings and good for ventilation success (Table 4.7).

Table 4.7. Anaesthetists' test – retest agreement (kappas) (n=22)

Rater	Overall success rating 1-5	Overall Success rating >3	Ventilation success (yes/no)	LMA handling rating 1-5
1	-*	0.54	0.65	-*
2	0.24	0.42	0.70	-*

* Kappa statistics could not be computed due to asymmetry of the result 2-way tables.

4.3.4. CLINICAL PRACTICE

Eighty-one participants (64.3%) returned one or more of their assessment forms and/or their clinical practice self-assessment questionnaire.

ARICM training

Data were gathered on the number of times that the participants had practiced on a mannequin as part of their ARICM module either before or after the baseline assessment and excluding intervention training (Table 4.8). The intervention and control group did not differ in the mean number of LMA placements practiced as part of ARICM.

Table 4.8. ARICM mannequin practice received

No of times of ARICM mannequin practice	Study group	N	Mean	SD
	Intervention	40	1.35	0.921
	Control	40	1.35	0.802
	Missing	1		

Intervention training

Data were gathered on the number of times that the intervention group participants (N=40) practiced on a mannequin as part of the study intervention, following the baseline assessment. Participants practiced LMA placement a mean ± SD (range) of 1.78 ± 0.70 (1-4) times in addition to their ARICM training.

4.3.4.1. INSTRUCTOR ASSESSMENTS

Eighty-eight percent (88%) of the participants achieved LMA placement on their first patient that led to successful ventilation. No statistically significant differences were found between the groups in the instructors' Likert-type ratings of overall LMA placement success (Table 4.9).

No significant differences were found between the groups in achieving successful ventilation in real patients and the trend was in the direction of more success among controls (Table 4.9). With the exception of LMA handling on the second patient, no statistically significant differences were found between the groups in the instructors' Likert-type ratings of LMA handling during the insertion, achieving effective ventilation in ≤40sec or achieving ventilation in at 1st insertion attempt (Table 4.9).

Table 4.9. The instructor outcome measures during clinical practice

		Intervention (I)	Control (C)	Test statistic	P-value
1st Patient	Overall success mean rating	3.34 (n=35)	3.55 (n=29)	$\bar{I} - \bar{C}$ -0.21	0.168*
	Overall success rating > 3	37.1 % (n=35)	48.3% (n=29)	X^2 0.806	0.369
	Successful ventilation (%)	82.9% (n=35)	96.6% (n=29)	Fisher's exact test	0.116
	LMA handling (mean rating)	3.23 (n=35)	3.40 (n=30)	$\bar{I} - \bar{C}$ -0.17	0.168*
	Time to task ≤40 sec (%)	74.3% (n=35)	72.4% (n=29)	X^2 0.028	0.866
	Success at 1st attempt (%)	66.7% (n=33)	74.1% (n=27)	X^2 0.388	0.533
2nd Patient	Overall success mean rating	3.69 (n=36)	3.69 (n=36)	$\bar{I} - \bar{C}$ 0.00	0.500*
	Overall success rating > 3	61.1% (n=36)	66.7% (n=36)	X^2 0.241	0.624
	Successful ventilation (%)	89.2% (n=37)	94.4% (n=36)	Fisher's exact test	0.674
	LMA handling (mean rating)	3.89 (n=36)	3.56 (n=36)	$\bar{I} - \bar{C}$ 0.33	0.031*
	Time to task ≤40 sec (%)	86.5% (n=37)	69.4% (n=36)	X^2 3.097	0.078
	Success at 1st attempt (%)	75.0% (n=36)	78.1% (n=32)	X^2 0.092	0.762
3rd Patient	Overall success mean rating	3.89 (n=35)	3.82 (n=33)	$\bar{I} - \bar{C}$ 0.07	0.418*
	Overall success rating > 3	71.4% (n=35)	66.7% (n=33)	X^2 0.180	0.671
	Successful ventilation (%)	91.7% (n=36)	82.9% (n=35)	Fisher's exact test	0.307
	LMA handling (mean rating)	3.89 (n=36)	3.76 (n=34)	$\bar{I} - \bar{C}$ 0.13	0.285*
	Time to task ≤40 sec (%)	80.6% (n=36)	65.7% (n=35)	X^2 1.994	0.158
	Success at 1st attempt (%)	81.3% (n=32)	64.5% (n=31)	X^2 2.238	0.135

*P-value obtained using Fisher's randomisation t-test

Table 4.9. The instructor outcome measures during clinical practice (continued)

		Intervention (I)	Control (C)	Test statistic	P-value
4th Patient	Overall success mean rating	3.68 (n=34)	3.70 (n=33)	$\bar{I} - \bar{C}$ 0.02	0.500*
	Overall success rating > 3	58.8% (n=34)	63.6% (n=33)	X^2 0.163	0.686
	Successful ventilation (%)	85.7% (n=35)	93.9% (n=33)	Fisher's exact test	0.429
	LMA handling (mean rating)	3.65 (n=34)	3.73 (n=33)	$\bar{I} - \bar{C}$ -0.08	0.399*
	Time to task ≤40 sec (%)	85.7% (n=35)	84.4% (n=32)	X^2 0.024	0.878
	Success at 1st attempt (%)	79.3% (n=29)	84.8% (n=33)	X^2 0.324	0.569

*P-value obtained using Fisher's randomisation t-test

The main reasons for failed LMA placement attempts included inadequate seal, failure of the participant to position the LMA and patient-related factors such as difficult airways, light anaesthesia or edentulous patients (Table 4.10).

Table 4.10. Reasons of LMA placement failure reported by the participants' instructors

	1st Patient		2nd Patient		3rd Patient		4th Patient		Total	
Complication	I	C	I	C	I	C	I	C	I	C
Inadequate seal	1			1		3	1		2	4
Difficult patient	2		2		2	1			6	1
Participant unable to pass LMA	2		1				1		4	0
Participant unable to pass LMA past the tongue	1					1	1	1	2	2
LMA positioned incorrectly	1	1		1	1				2	2
Participant abandoned attempt	1								1	0
Run out of time					1				1	0
Not reported			2				1		3	0
Total	8	1	5	2	4	5	4	1	21	9

I = Intervention group, C= control group

4.3.4.2. PARTICIPANT SELF-ASSESSMENTS

No statistically significant differences were found between the groups in their Likert-type self-ratings of overall performance (Table 4.11).

Table 4.11. Participant self- ratings of overall LMA placement success in patients

	Intervention mean rating	Control mean rating	Mean rating difference	P-value*
1st Patient	2.69 (n=32)	2.87 (n=31)	-0.18	0.219
2nd Patient	3.03 (n=34)	3.11 (n=38)	-0.08	0.404
3rd Patient	3.35 (n=34)	3.46 (n=37)	-0.11	0.352
4th Patient	3.39 (n=33)	3.68 (n=37)	-0.29	0.120

*P-value obtained using Fisher's randomisation t-test

No statistically significant differences were found between the groups in their Likert-type self-ratings of LMA handling during the insertion (Table 4.12).

Table 4.12. Participant self-ratings of handling the LMA during insertion in patients

	Intervention mean rating	Control mean rating	Mean rating difference	P-value*
1st Patient	2.81 (n=32)	2.87 (n=31)	-0.06	0.427
2nd Patient	3.15 (n=34)	3.16 (n=38)	-0.01	0.500
3rd Patient	3.32 (n=34)	3.38 (n=37)	-0.06	0.443
4th Patient	3.58 (n=33)	3.54 (n=37)	0.04	0.487

*P-value obtained using Fisher's randomisation t-test

When comparing the participants' self ratings of overall LMA placement success to the same ratings by their instructor, participants tended to underrate their first two LMA placements (Fig. 4.4). The agreement between the participants and their instructors' ratings in terms of the participants' overall LMA placement success was non-significantly poor (Table 4.13). Their agreement was non-significantly poor to fair in terms of having assigned an above average (>3) overall success rating.

Table 4.13. Instructor and participant agreement (kappa) during clinical practice – Overall success

	Kappa	
	Overall success rating 1-5	**Overall success >3**
1st Patient	-*	0.12†
2nd Patient	-0.01†	0.09†
3rd Patient	-*	0.33
4th Patient	0.15†	0.41

* Kappa statistics could not be computed due to asymmetry of the result 2-way tables.
† Statistically non-significant

The agreement between the participants and their instructors in terms of the participants' handling of the LMA during the insertion was poor (Table 4.14).

Table 4.14. Instructor and participant agreement (kappa) during clinical practice – LMA handling

	Kappa **LMA handling 1-5**
1st Patient	-*
2nd Patient	-*
3rd Patient	0.20
4th Patient	-*

* Kappa statistics could not be computed due to asymmetry of the result 2-way tables.

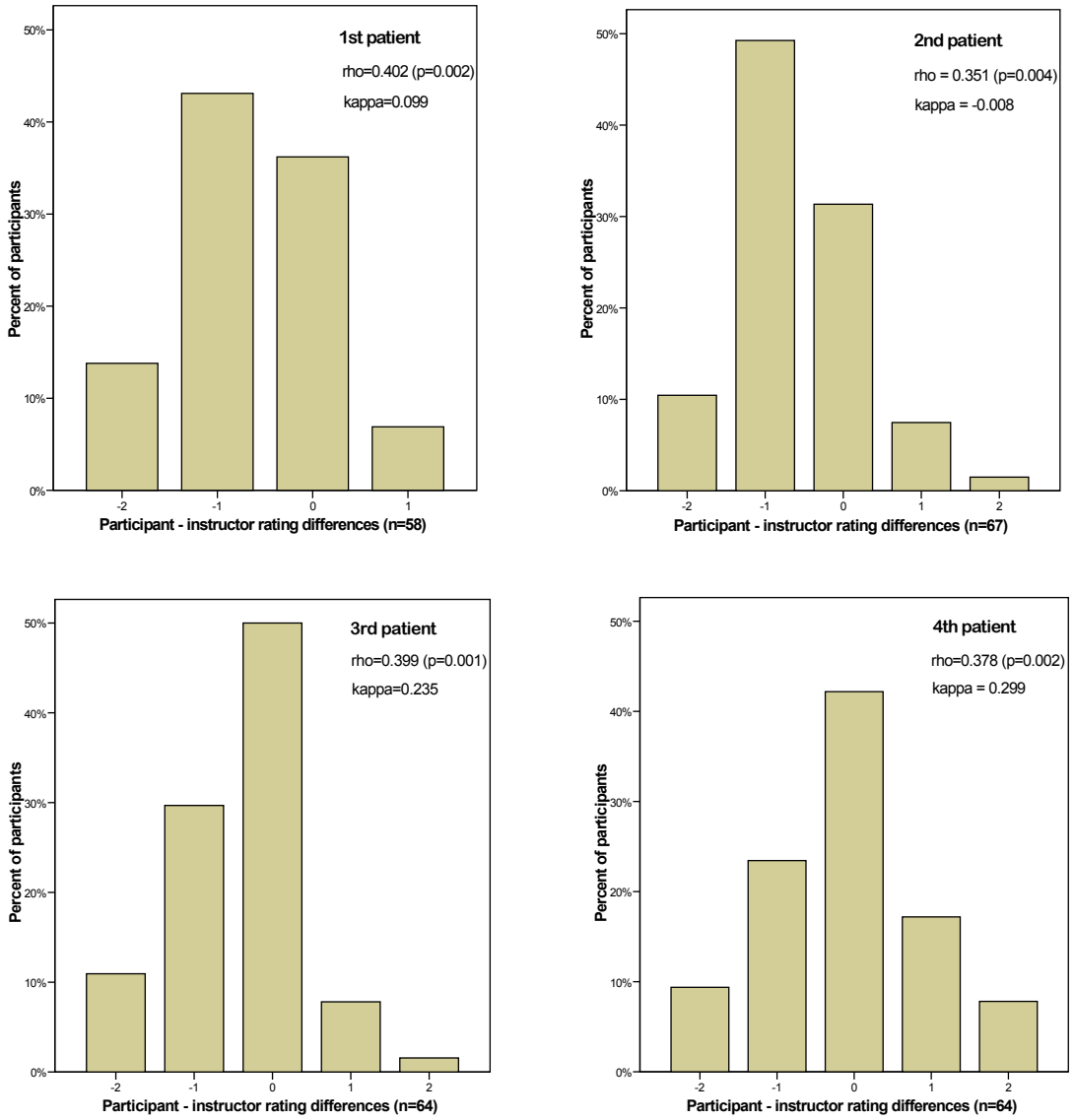

Figure 4.4. Differences between participant self-ratings and instructor ratings for overall success of LMA placement (difference = participant rating – instructor rating)

4.3.5. PARTICIPANT FEEDBACK

Seventy-two participants (57%) returned their self-assessment clinical practice questionnaires. 69.4% of the respondents 'agreed' or 'strongly agreed' that their overall mannequin training had been helpful, 14.1% 'disagreed or strongly disagreed', 9.7% were 'undecided' and 5.6% indicated that they had not had any mannequin practice' (Table 4.15). The latter had been in the control group.

Table 4.15. Participants' perceptions of their mannequin training

Mannequin training was helpful	**Intervention (n=35)**	**Control (n=37)**
Agree	65.7%	73.0%
Disagree	22.9%	8.1%
Undecided	11.4%	8.1%
N/A	0%	10.8%

There was no statistically significant difference found between the two groups in terms of the number of participants that had found their mannequin training helpful (X^2= 2.263, df=1, p=0.132).

Fifty-seven of the respondents, 30 intervention group and 27 control group participants, opted to include some further qualitative feedback. The main themes that were mentioned were:

- Twenty-four respondents (42.1%) felt that the mannequins mimicked reality poorly leading to a major limitation in their training value.

'Was useful to see the equipment etc used and get used to handling it, but was very different when dealing with a real patient and real environment'. (control group)

'Placing an airway on a stiff rubber mannequin is nothing like doing it on a patient so was of no use in preparing for performing the procedure on a patient'. (intervention group)

'It familiarised me with the LMA, however, it is much different on a patient since their tongue can easily fall back'. (control group)

Two respondents elaborated further on the subject of how clinical practice differs from mannequin practice by adding the themes of fear of hurting the patient and the clinical variation seen in real patients.

'*However, simulator can't prepare you for difference in patients- they feel different to dummies, have different sized mouths, you are worried not to hurt them, etc.'* (intervention group)

'Whilst extra training on how to insert an LMA may affect a student's performance during insertion, I think that it is more the 'type' of patient you have to insert on in, is the biggest limiting factor for learning, for example, the age and whether they have any teeth or not makes a big difference'. (control group)

- Nineteen respondents (33.3%) reported that mannequin training had helped them to learn the basic technique required in order to place and LMA before approaching any patients.

'The simulator training was useful to obtain a basic idea of the order of events when inserting an LMA'. (intervention group)

'Need to learn basics before shoving a big bit of plastic down someone's throat'. (control group)

- Eight respondents (14%) reported that mannequin training had made them more confident.

'I am a particularly nervous student when it comes to practical procedures therefore this extra opportunity to practice before a real situation was greatly appreciated!' (intervention group)

'I found my ARICM mannequin training very helpful because it gave me the confidence that I knew the mechanism of putting in an LMA before being let loose in theatre.' (control group)

- Six respondents (10.5%) commented that being taught on real patients was superior to mannequin training.

'The training was useful in knowing how to perform the LMA but was different in practice. It would be more useful to be shown on a real person although not very practical'.
(Intervention group)

'The simulator training on the dummy was helpful to a certain extent. I found it more useful being shown by a consultant on a real patient'. (Control group)

- Five respondents (8.7%) commented that there was a discrepancy between the technique they had been taught during their mannequin training and the methods used in the hospitals.

'*At least when consultants ask if you have done it before you can say yes so they will let you have a go. When consultants tell you how to use them they just say 'shove it in' which isn't very helpful'.* (Intervention group)

The methods used in hospitals are different to the method taught in the training'. (Intervention group)

4.4. DISCUSSION

Brief summary of findings

The purpose of this study was to investigate whether a short period of additional mannequin training, added to a very basic exposure of LMA placement on mannequins, would increase medical students' LMA placement success during their regular clinical practice with real patients. The participants who received the additional mannequin training session achieved similar success rates in placing LMAs to those who received only one brief mannequin training session across all outcomes studied. A statistically significant difference in favour of the intervention group was found in the instructors' Likert-type ratings of LMA handling during the insertion on the second patient, however, this was not corroborated by any other result and may be the result of multiple statistical testing.

Participant self-ratings of performance did not differ between the two groups. Overall participants tended to underrate their performance for the first two LMA placements and agreement between participant ratings and their instructors' ratings was generally poor to fair. Participants' perceptions of the value of their mannequin training did not differ significantly between the intervention and the control group. According to their qualitative feedback, mannequins were poor at mimicking reality, limiting their training value to familiarising themselves with the LMA device, basic technique and acquiring some confidence prior to being faced with the real situation.

The current study findings suggest that merely increasing the 'dose' of mannequin practice yields no additional benefit. Our results broadly corroborate the only other controlled study of LMA placement training that assessed effectiveness of different 'doses' of mannequin training[106]. In this study the performance of a group who had practiced 10 times on the mannequin was not statistically significantly different in terms of placement grades from a group who had practiced 5 times. Nevertheless, in contrast to our findings the trend was in the direction of improved performance with more opportunity to practice.

Limitations

Ethical considerations prevented including a no practice group in this study as mannequin practice in LMA placement is a set part of the 4th year medical student curriculum. As part of their ARICM module all students are required to undergo mannequin practice training prior to practicing on patients and depriving students of this training was not deemed acceptable.

A large number of different instructors took part in the assessment and they had no specific training in evaluating proficiency for this study. The baseline data showed that the rater agreement in five different pairs of assessors using the Likert-type scales of overall LMA placement success and LMA handling during the insertion ranged from poor to fair although agreement between the different groups of assessors ranged from moderate to very good in the assessment of adequacy of ventilation (ventilation success) by direct visualisation of the mannequin. The baseline intra-rater agreement with regard to the Likert-type scales of overall LMA placement success for the first block was fair. The test-retest agreement was good with regard to ventilation success. Fiberscopy after insertion of the LMA is recommended as a useful way of assessing the mask position as the LMA can allow adequate ventilation even if sub-optimally placed[110]. However, the present field situation involved participants practicing simultaneously in multiple sites which made the use of fiberscopy impractical and the instructors were not known beforehand.

A significant number of participants were lost to follow-up despite email reminders and eligibility in a prize draw for vouchers. Low response rates can lower the chance of getting similar groups in terms of key characteristics. However, these losses did not seem to have imbalanced the characteristics of the participants from baseline to follow-up. A similar number of participants dropped out from both the intervention and control groups.

Incomplete data among both participants and their instructors also limited the analysis. As a result in many cases data on individual attempts were missing completely or insufficient information was provided on reasons for failure and exact time to insertion. Prior to the study it was thought that LMA placement was an invariable component of ARICM training. It

transpired that some students do not get to practice LMA placement four times or more, and in two cases students had no chance to practice LMA placement during their ARICM hospital placement. Furthermore, there were cases were students lost the forms or forgot to take them to the operating room.

Consideration of possible mechanisms and explanation

Davies et al[103] reported 100% success in their participants' first patient in terms of LMA insertion, 82% success for the second patient, above 90% subsequently and an overall success rate of 94%. Their first patient success rates are higher than in this study. However, the choice of patient where trainees were tested on was controlled i.e. ASA 1 patients with no loose teeth or crowns and their training intervention included additional instruction using videotapes. Overall, their success rates seem to be comparable to the current study. Roberts et al[111] found a 98% LMA insertion success rate in their 'mannequin training only' group. This success rate was comparable to this study's 'basic mannequin training' control group. The current study findings suggest that merely increasing the 'dose' of mannequin practice past an initial brief session leads to no additional benefit, at least when this involves low-fidelity mannequins. Three (non-exclusive) reasons might explain the null results of this study.

Firstly, the training offered may have been sub-optimal. However, in a recent systematic review repetitive practice and providing feedback were the top two features acknowledged as important for effective learning in medical simulations[16] and these tenets of good practice were followed in the intervention training. Conversely, a recent systematic review of comparative studies of clinical skills training that focused on intubation, venous cannulation

and central venous line insertion concluded that the addition of simulators, including mannequins, to a traditional course was not supported by study results[82].

A second possibility is that the mannequins were too basic. They consequently failed to provide some other important features identified in Issenberg et al's review; providing a range of difficulty levels, capturing clinical variation and providing high degrees of realism[16]. Many of the participants drew attention to the rather low fidelity of the simulators, a notion that has been previously expressed[112, 113]. Furthermore, simulators cannot mimic the existential experience of clinical practice, especially in the atmosphere of an acute setting such as an operating theatre. The improvement observed over only four patients in real practice settings adds credence to this explanation.

The third explanation rests on the observation that LMA placement is a relatively simple technical skill to acquire[114] – this is in large measure the reason behind the introduction and rapid dissemination of this method of airway management[102]. This observation leads to a hypothesis for further testing: the need for simulation training increases with the degree of difficulty of the task. Below a certain difficulty threshold, it may be more cost-effective to train 'at the bedside', with minimal possible risk to patients. While few would doubt the validity of simulation training for technically demanding procedures (e.g. key-hole surgery), according to this hypothesis it would be unnecessary and wasteful to insist that simulation training be used in all circumstances.

Thus, a theory might be that the effectiveness of well taught *in vitro* training of practical skills depends on two factors: 1) The extent to which the *in vitro* simulation mimics the skill required. Note that the crucial point here is the skill, not the simulator *per se* – the motor skills needed for tying surgical knots or performing three-dimensional key-hole surgery do not require replication of the entire setting, merely the activity. In other settings 'willing suspension of disbelief' may sufficiently replicate real settings. However, in other settings the 'feel and texture' of the simulation may be important. 2) The degree of difficulty of the procedure. This hypothesis needs to be tested on a wide range of technical tasks and this would require the generation of a difficulty scale. However, the hypothesis, if confirmed, would have large implications for health care allowing expensive training resources to be focused where they provide the most benefit.

Summary of clinical and research implications

A substantial amount of the participants' comments indicated that the lack of realism of the mannequins limited their training effect. These comments seem to complement the findings of this study and carry four main implications. Firstly, the basic mannequin airway significantly differs from the human airway in important aspects and the clinical context is far removed from the training environment. The users should be made aware by their instructors of these aspects and of clinical variations encountered in real life. Secondly, future research should seek to compare the basic mannequins currently used for training to training on high-fidelity mannequins that provide a range of difficulty levels, and to assess whether this latter training would be more effective in preparing students for real life practice. Thirdly, great care should be exercised in interpreting self-rated assessments and they should not be used as a surrogate for observations of clinical proficiency. Fourthly, future research should test the

hypothesis that it may be more cost-effective in terms of training outcomes to train under supervision ‘at the bedside’ for technical tasks such as LMA placement that have a low degree of difficulty at the initial stages of learning.

CHAPTER 5

THE EFFICACY OF THE MANAGEMENT OF LIFE-THREATENING ILLNESS SIMULATION COURSE: A RANDOMISED CONTROLLED TRIAL

5.1. INTRODUCTION

In the UK, postgraduate courses teaching basic and advanced resuscitation skills deal predominantly with the management of a patient in the period after a catastrophic event such as cardiac arrest[115]. However, some patients suffering catastrophic deterioration have premonitory symptoms and might have improved outcomes with earlier recognition and appropriate management[116]. Hospital-wide approaches to the management of patients at risk by early recognition of deterioration and early resuscitation have been developed to reduce the number of unexpected deaths, cardiac arrests and unplanned intensive care unit admissions[117, 118]. In addition, there are training programmes, such as ALERT[115] that teach a pre-emptive approach to critical illness at postgraduate level.

A priority for medical education is therefore to supply training that ensures students develop the skills needed to recognise and treat critical events[119]. However, caring for acutely ill patients is a difficult area to teach trainees as patient safety concerns and the unpredictable occurrence of emergencies restrict training opportunities[2, 120]. In response, simulations are increasingly used to teach practical skills and team working in the healthcare context[1, 2, 22].

To address the need to teach the skills needed to recognise and prevent critical events at undergraduate level, the department of Anaesthesia of The University of Birmingham has developed a simulation course named Management Of Life Threatening Illness (MOLTI). This course is aimed at 5th year undergraduate medical students. During MOLTI students act out the management of potentially life threatening scenarios using a mannequin, X-ray and patient chart laminas, a box of potentially relevant medical devices and a special e-learning tool that generates oxymeter sounds that can be altered at will by the course instructor. There are six scenarios available and students can assume the role of the physician or nurse or be an observer. Each scenario has a predefined sequence of when and how the simulated 'crisis' evolves. A description of the scenarios can be found at Appendix 5.1. The MOLTI course aims to help undergraduate medical students to 1)understand the importance of reversing adverse physiological trends quickly, 2)understand the importance of providing Oxygen, IV access and fluids and 3)practice immediate assessment, monitoring and treatment based on a simplified version of the ALERT framework (Fig. 5.1).

- **A**irway
- **B**reathing
- **C**irculation
- **D**isability
- **E**xposure
- Basic treatment- Oxygen, IV fluids
- Call for help if necessary
- Full examination
- Timely investigations
- Periodic reassessment of the patient

Figure 5.1. The MOLTI framework

The MOLTI course is delivered in a collaborative learning format i.e. participants are trained in groups. Collaborative protocols have been shown to reduce required instructor time and resources by half and provide observational learning opportunities that compensate for hands-on practice efficiently and effectively, as predicted by social learning theory[24]. The MOLTI course, which is not mandatory, has been run since 2005 as part of the University of Birmingham MBChB Year 5 Surgery phase. It has been well received by students and instructors and its use is increasing. However, there has been no research to assess its efficacy.

5.1.1. RELATED STUDIES

A literature review was performed to investigate the evidence regarding the effectiveness of scenario-based computer-controlled mannequin training. Studies where included if they had a) investigated the effectiveness of computer-controlled mannequin training, b) the training was related to management of life-threatening illness, c) the participants were undergraduate students and d) quantitative outcome measures had been used. The ideal study would compare outcomes in an intervention group with outcomes in a randomly generated control group. However, given the dearth of RCTs assessing students' performance, studies lacking a control group or based on participants' reactions/self-ratings were also included.

The following studies were identified:

Controlled studies in vitro

Alinier et al[121] investigated the effectiveness of scenario-based simulation training on second-year nursing students' clinical skills in an RCT (n=99). The trial compared traditional training which did not include scenario-based simulation to traditional training including two

3-hour simulation sessions on the SimMan. The participants were assessed using an OSCE and checklists prior to the training and at least five weeks after the simulation sessions of the intervention group. The intervention group achieved a statistically significantly higher mean test score compared to the control group. Nevertheless, the simulation-based training did not have a statistically significant effect on participants' self-reported confidence about 'working in a highly technological environment'. The methods of randomisation and allocation were not reported and it was not specified whether the assessors were blinded to group allocation.

In another RCT Steadman et al[122] compared full-scale simulation training using the Medical Education Technologies Inc (METI) patient simulator to interactive problem-based learning (PBL) for teaching fourth-year medical students (n=31) acute care assessment and management skills. Participants were randomised to the simulation or the PBL group using a computer randomisation program. All participants underwent a simulator-based initial assessment and both groups learned about dyspnoea. In order to equalise simulator education time, both groups learned about abdominal pain with the PBL group using the simulator and the simulator group using the PBL format. Subsequently they were all tested on a different dyspnoea scenario using the simulator. Each participant was assessed by two blinded assessors using a standardised checklist. The simulation group achieved a statistically significantly higher mean test score compared to the PBL group in the final assessment.

In a randomised trial Morgan et al[123] investigated the effect of high-fidelity simulator training on final-year medical students' performance on managing a critical event and written examination marks (n=144). The trial compared video-based training using videotaped demonstration by a faculty member appropriately managing a simulator scenario to a

simulator session consisting of a preprogrammed scenario supervised by a faculty member or senior anaesthesia resident. Students were randomly allocated using computer-generated, randomly selected numbered sealed envelopes. The participants were assessed pre- and post-test on the same day using a standardised checklist on the same critical event scenario and by a written examination 2, 16, or 30 days following training. There was a significant improvement in post test scores over pretest scores for both groups. However, there were no statistically significant differences in checklist scores or final examination marks between the simulation and video training groups. The authors did not state whether assessors were blinded. Student opinions indicated that the simulator training sessions were more enjoyable and considered more valuable than the video teaching sessions.

Uncontrolled studies in vitro

MacDowall[124] investigated the perceived confidence and competence of final-year medical students (n=23) in the assessment and treatment of the acutely ill patient by administering a questionnaire before and after training in the management of the acutely ill patient on the SimMan, a computer-controlled mannequin. The students used five-point scales to rate statements regarding their confidence and ability. A statistically significant improvement was found in all questions.

Weller et al[120] conducted a study to assess the ability of undergraduate medical students (n_1=45 4th year students, n_2=26 6th year students) to manage medical emergencies and the educational value of simulation using the SimMan. Following an initial familiarisation period, participants attended a 3-hour simulation-based workshop where they worked in teams in one of three post-operative shock scenarios. Each team completed the same scenario twice

(baseline and repeat). Participants also observed peers, received feedback following simulations and participated in a facilitated discussion to develop a systematic approach to the shocked patient. The first 5-minute period of each scenario was standardised and used to compare baseline and repeat scenario performance. Videos of the baseline and repeat simulations were randomised and assessed independently by two independent assessors using a global performance scores (based on 5-point scales of systematic approach to the problem, leadership and division of tasks between team members) and a checklist. The assessors were blind to the year level of the participants and the order of the scenarios. A significant improvement in global and checklist scores was found from baseline to repeat but the assessment may have been biased by the learning effects of repetition.

In another study by Weller[125], 33 fourth-year medical students attended a simulation workshop on management of medical emergencies using the SimMan and answered a post-course questionnaire on their perceptions regarding learning outcomes. They also scored their level of competency on the training material using a 5-point scale before and after the session. A statistically significant mean increase in competency (SD) of 1.11 ($p<0.0001$) was reported. Participants rated the workshop highly and identified teamwork skills, learning how to approach a problem better and how to apply theoretical knowledge as key learning points. However, this study did not include any other form of assessment.

Morgan & Cleave-Hogg[126] conducted a study to determine the reliability of assessments of medical students using the Anaesthesia Simulator and to elicit student opinion (n=24). All participants underwent a scenario that reflected objectives of the anaesthesia rotation. The

simulator experience was rated very positively by students using a 5-point scale for learning experience (Mean = 4.6) and appropriate content (Mean = 4.4).

Bearnson & Wiker[127] conducted an exploratory, descriptive study to explore the benefits and limitations of using a human patient simulator to teach medication administration to first-year nursing students. Each student underwent a 2-hour simulation session using three scenarios (a healthy adult, a middle-aged obese woman with respiratory problems and an elderly hypertensive male). Participants rated positively their simulator experience using a 4-point scale for knowledge on medication side-effects (Mean =3.13), knowledge of differences in patients' responses (Mean=3.31), ability to administer medications safely (Mean=3.06) and increased confidence in medication administration skills (Mean=3.00).

However, as commented by Alinier et al[121] most studies on patient simulators have low statistical power or rely solely on participant feedback[124-127]. None of the studies reviewed used *in vivo* evaluations.

5.1.2. STUDY AIMS

The aims of this study were a) to assess the efficacy of the MOLTI course using the SimMan b) to compare undergraduate self-assessment of success in the management of life threatening illness with objective assessment by a third party and c) to explore participant perceptions of the course.

5.2. METHODS

Clinical context

During the academic year 2007-2008, all Year 5 University of Birmingham medical students undertook their MBChB clinical training in one of six groups that alternated between six phases, each phase lasting six weeks. As part of their Surgery phase, all students had to successfully complete their Immediate Life Support (ILS) training and some were also offered additional MOLTI or ALERT training, depending what was available in their respective hospital.

Recruitment

The study was approved by the Medical School Education Unit, The University of Birmingham and consent was obtained from all the participants. Eligible participants were all Year 5 University of Birmingham medical students who had not done the MOLTI course previous to this study.

The study was first advertised to the whole Year 5 through an email invitation in June 2007. Additional email invitations were sent to each of the six individual groups prior to commencing their Surgery phase and a short introductory talk was given during one of their academic in days. Students were offered a study information sheet and were asked to register their interest by providing their electronic address. Participants were recruited from July 2007 to February 2008.

Students who agreed to participate by providing their consent via email were provided with a set of dates and were re-sent the study information sheet and further information regarding the

venue. The study was conducted in 20 sessions and each participant took part in two of these sessions, which were one week apart (Fig. 5.2).

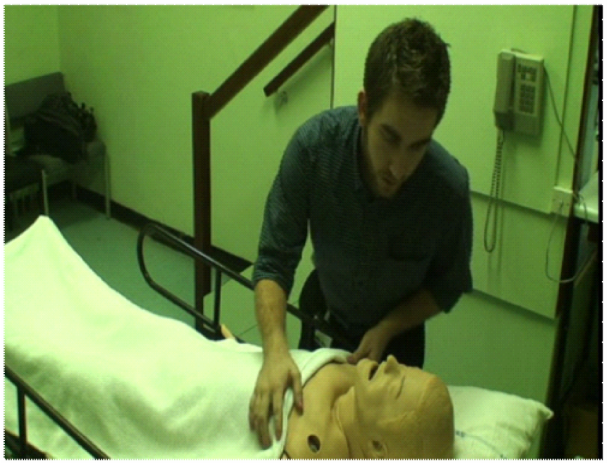

Figure 5.2. Participant undertaking a MOLTI scenario

Baseline assessment and randomisation

On the agreed starting date, participants undertook a baseline assessment which took place in the Department of Anaesthesia and Intensive Care at the Queen Elizabeth Hospital. All participants were initially asked to fill in a brief baseline questionnaire asking about their demographics, their previous training experiences and their perceived confidence in providing the right basic management to a deteriorating patient (Appendix 5.2). They were also given an introductory talk reiterating what would be asked of them during the study. They were shown the SimMan and were given an explanation of its capabilities.

Each participant undertook a baseline assessment on a six-minute tension pneumothorax scenario. The scenario concluded either with resolution of the crisis or at the discretion of the simulation instructor at the end of six minutes. Subsequently each participant was asked to

open an envelope containing their group allocation. Block randomization and sequentially numbered sealed opaque envelopes were used for the allocation. Participants were randomised to receive either the MOLTI training on the SimMan (intervention group) or no training (control group). A random-number table and blocks of four were used to generate the random allocation sequence as described by Altman & Bland[107]. The researcher administering the allocation (Celia A. Brown) was independent of the recruitment process and did not have any knowledge of the participants' baseline characteristics.

The participants' performance was recorded using the SimMan software (Appendix 5.3) and a video camera. The collected material was rated independently by two anaesthesiologists using a checklist specifically devised for this study (Appendix 5.4), a global rating tool which was an adapted version of a previous rating tool developed by Weller et al[128] (Appendix 5.5), and an overall 'pass/fail' mark. The participants were also asked to rate their overall performance on a 5-point scale which was identical to the 'overall performance' scale of the global rating tool used by the assessor. A flow chart of the assessments used in this RCT is presented in Fig. 5.3.

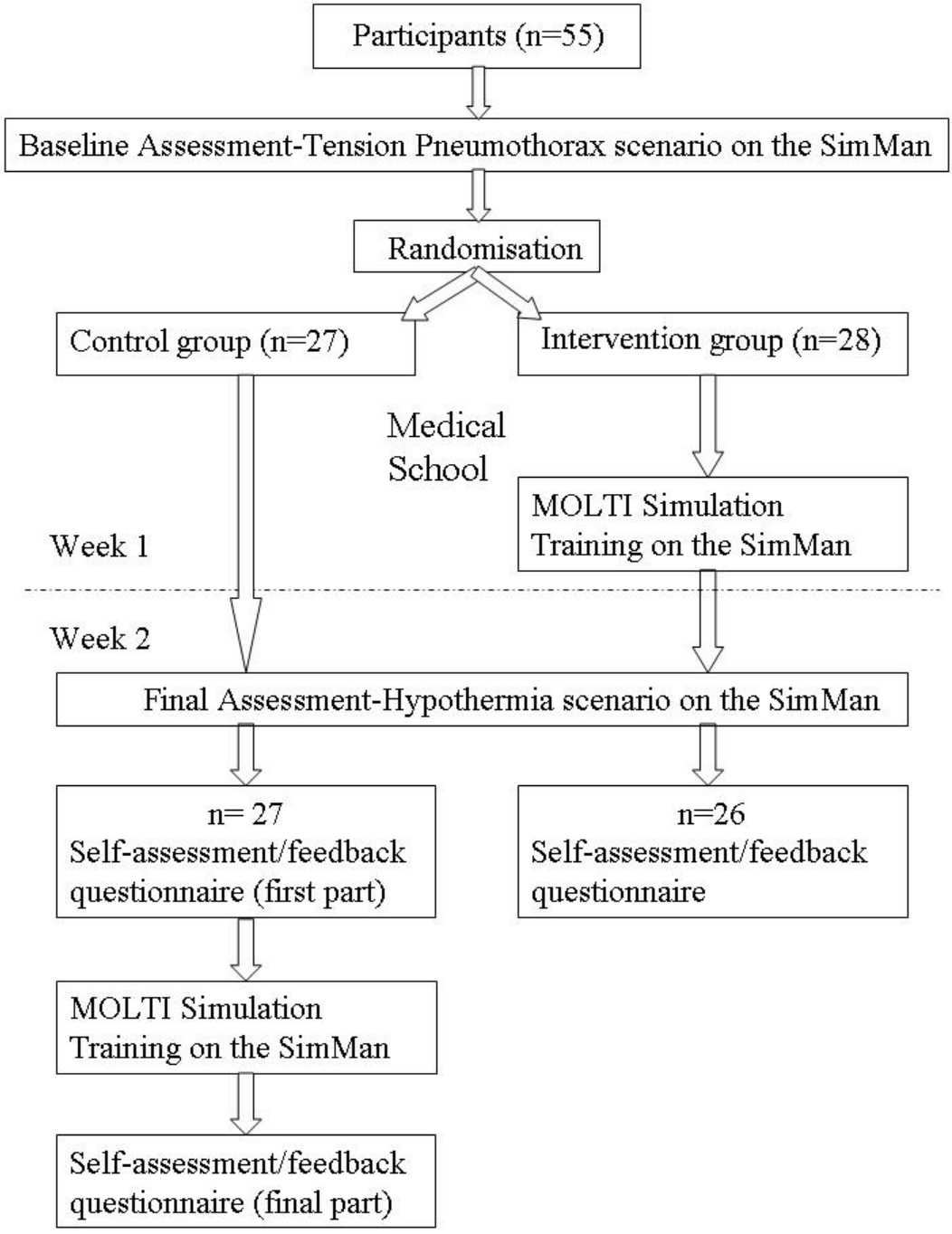

Figure 5.3. Flow chart of the RCT's assessment process

Intervention

Participants in the intervention group received half an hour of group training which consisted of feedback on their tension pneumothorax performance and going through additional MOLTI scenarios on the SimMan. The training objectives of the MOLTI course were reiterated throughout the session while additional individual management elements for each scenario were introduced. Participants in the control group received no MOLTI training following their baseline assessment and their colleagues in the intervention group were implicitly instructed not to reveal the content of the MOLTI course.

Final assessment

All participants were asked to return for a final assessment one week later. Each participant undertook an assessment on a six-minute hypothermia scenario. Again, their performance was recorded using the SimMan software (Appendix 5.6) and a video camera, and the collected material was rated independently by two anaesthesiologists using a second checklist (Appendix 5.7), the adapted global rating tool[128] and an overall 'pass/fail' mark. The anaesthesiologists were blinded to the participants' group allocation and were not told which scenario was the baseline and which was the final. The participants were also asked to rate their overall performance on a 5-point scale which was identical to the 'overall performance' scale of the global rating tool used by the assessors. Following this final assessment, constructive feedback was given to all participants by the instructor. The participants in the control group were then offered the MOLTI training. All participants were asked to fill in a feedback questionnaire which included items of previous questionnaires by Holzman et al[129] and MacDowall[124] (Appendix 5.8). All participants were asked not to reveal the content of the MOLTI course to any of their colleagues who had not yet participated in the study.

Sample size

The number of participants for the study was decided using a binomial power calculation. The main study sample size was calculated using the UCLA Department of Statistics power calculator. A binomial distribution 2-sample arcsine approximation was used. The probability of students having acceptable success in taking the necessary actions in order to avert catastrophic 'patient' deterioration (exhibited as an overall global rating ≥3) was set at 0.75 for population 1 (control group) and 0.99 for population 2 (intervention group). Probabilities were derived from expert opinion. Significance level was set at 0.05. The power was set at 0.9 (90%), for a one-sided test. A one-sided test was used as it was hypothesised that training would not result to poorer management skills than no training. The number of participants in each group was calculated to be 24. To compensate for drop out/ failure to complete, the planned number of participants was 30 per group.

Analysis

Statistical analysis was performed using SPSS 12.0.1 and R 2.6.2 for Windows. The distribution of numerical data was determined using the Shapiro-Wilk test of normality. Fisher's randomization T-test was used to compare groups in terms of the assessor-awarded medical management, behaviour and overall performance ratings of the global rating tool and overall performance self-ratings. Pearson Chi-squared analyses were used to compare groups in terms of rater pass/fail marks and achieving overall performance global ratings ≥ 3. Participants were awarded a pass if *both* raters had awarded a pass mark. Fisher's Exact Test was used when expected counts were less than 5.

Due to technical failures, a small amount of the SimMan logs and a significant amount of the

video recordings were lost. The assessors rated each case using all available data. To generate participant checklist scores for each scenario, the average of the checklist ratings for the two raters was used. These scores were converted to percentages based on the resultant mean performance and the maximum score that could be obtained for a given simulation scenario similar to Boulet et al[130]. Independent samples T-tests were used to compare checklist percentage scores and times to administering O_2 and I.V. fluids. The Mann-Whitney U test was used to compare times to administering O_2 and I.V. fluids where data were not normally distributed. The analyses were repeated for the subgroup of participants for whom all data had been collected successfully both at baseline and final assessment.

Inter-rater agreement and rater-participant agreement were explored using the kappa statistic[109]. The participant's Likert-type confidence ratings before and after simulation training were analysed using the Wilcoxon signed-rank test. A p-value of <0.05 was considered statistically significant. The participants' quotes resulting from the clinical practice self-assessment questionnaire were categorised into common themes.

Pilot study

In May 2007 the study protocol was tested on 8 fifth-year participants who volunteered following an email invitation to the whole 2006-2007 cohort. As result, it was determined that no more than 12 students could be admitted to each study session. The data collected were used to pilot the checklists and the global rating tool used. Furthermore, part of these data was jointly reviewed by the two independent assessors of the study to establish common understanding of the rating process.

5.3. RESULTS

5.3.1. POPULATION

Fifty-five Year 5 medical students enrolled in the study and participated in the baseline assessment. Participant characteristics are shown in Table 5.1. Table 5.2 contains data on the participants' demographic aspects versus those of the overall MBChB Year 5 population. Data on nationality were used as data on ethnicity could not be obtained for the overall population. Ninety-six percent of these participants completed the study by undertaking by coming back for the second session (Fig. 5.4). The study population and the overall MBChB Year 5 population were similar in terms of mean age and nationality and their difference in terms of sex was small (Table 5.2).

Table 5.1. Participant characteristics at baseline

Study group	**Intervention**	**Control**
No of participants	28	27
Age mean (SD) years	24.7 (1.5)	24.2 (1.1)
Sex: Male	10 (35.7%)	7 (25.9%)
Female	18 (64.3%)	20 (74.1%)
Ethnicity: British White	16 (57.1%)	21 (77.8%)
British Asian	4 (14.3%)	3 (11.1%)
British other	3 (10.7%)	1 (3.7 %)
Non British	5 (17.9%)	2 (7.4%)
Prior training:		
Computer simulation	4 (14.3%)	4 (14.8%)
ALERT	3 (10.7%)	0 (0.0%)
ILS	1 (3.6%)	4 (14.8%)
None	22 (78.6%)	19 (70.4%)

Table 5.2. Participant characteristics versus overall MBChB Year 5 population characteristics

Group	**Study participants**	**Overall population**
No of participants	55	403
Age mean (SD) years	24.5 (1.3)	24.6 (1.8)
Sex: Male	17 (30.9%)	165 (40.9%)
Female	38 (69.1%)	238 (59.1%)
Nationality: British	48 (87.3%)	367 (91.1%)
Non British	7 (12.7%)	36 (8.9%)

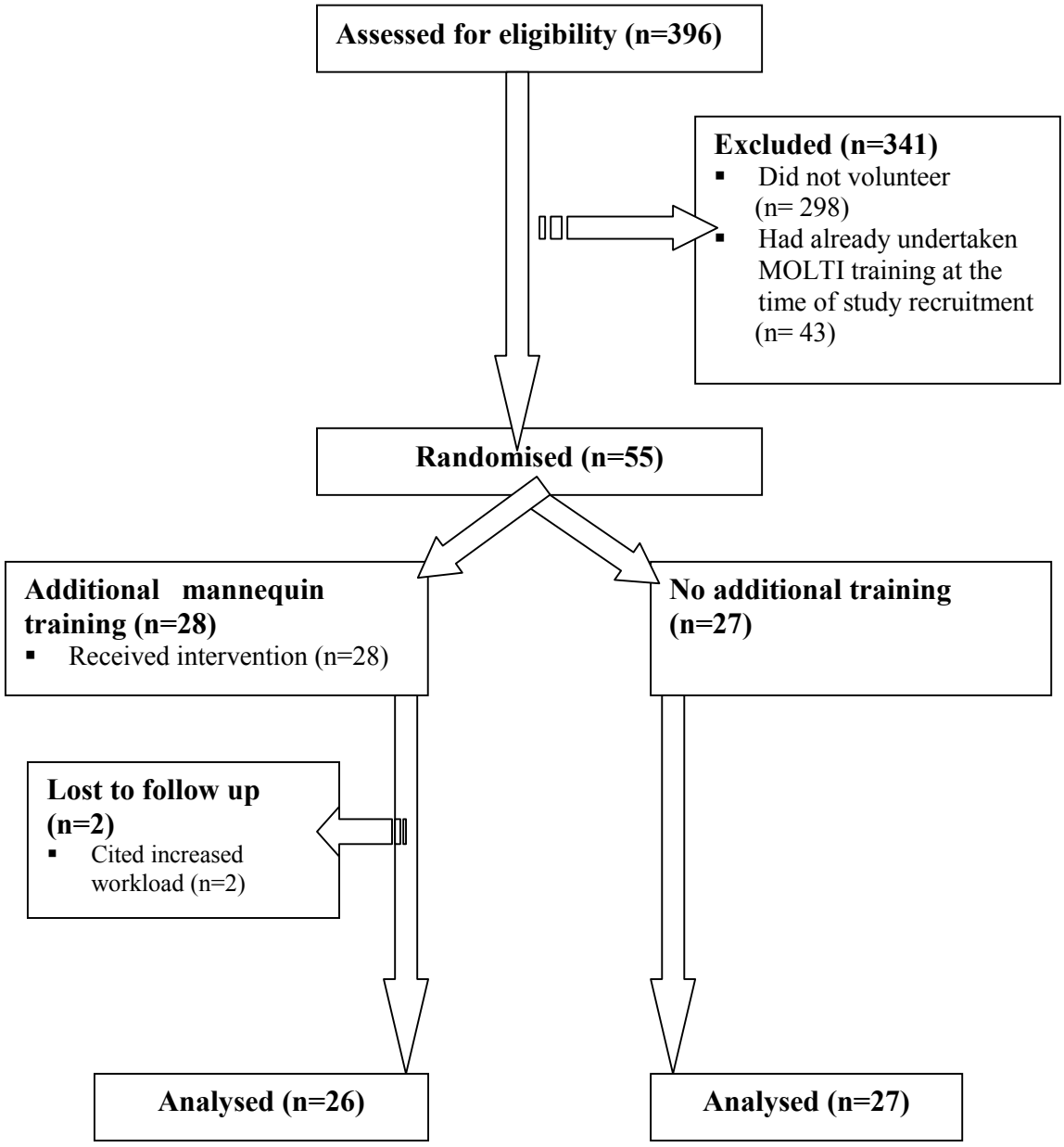

Figure 5.4. Participant flow chart following CONSORT scheme

5.3.2. PRIOR EXPERIENCE

All participants had received Basic Life Support training and two years of clinical training, including ARICM in their fourth year. Participants were asked about any other previous experiences with computer-based simulation and other training related to MOLTI, such as Acute Life-threatening Events - Recognition and Treatment (ALERT^{TM}) and Immediate Life Support (ILS). Five participants (5%) had previously undertaken ILS, three participants (11%) had undertaken ALERT and eight participants (15%) had previously engaged in some form of computer-based simulation (Table 5.1).

5.3.3. RATER ASSESSMENTS

The availability of video recording is shown in Table 5.3.

Table 5.3. Data availability

No of Videos retained	Intervention group	Control group
Baseline assessment	9 (32.1%)	8 (29.6%)
Final assessment	9 (43.6%)	15 (55.6%)

The effect of this loss was further investigated and it was observed that participants whose video record had been preserved were more likely to receive an overall 'pass' mark than the participants whose video record had been lost (Table 5.4).

Table 5.4. The distribution of Pass/Fail marks according to video availability

Video availability		Pass	Fail	X^2	P-value
Baseline Assessment	Yes	5	12	Fisher's exact test	0.020
	No	1	28		
Final assessment	Yes	8	16	1.837	0.175
	No	5	24		

5.3.4. BASELINE ASSESSMENT

All 55 participants underwent a baseline assessment which consisted of managing a MOLTI scenario of tension pneumothorax. The data for 46 participants (84%) were recorded. No statistically significant differences were found between groups at baseline, with the exception of medical management mean ratings where the control group was found to be statistically significantly better (Table 5.5). However, this was not corroborated by the other results and could be the result of multiple statistical testing.

Table 5.5. Baseline assessment

	All participants				Participants with both baseline and final assessment data			
	I* (n=24)	**C* (n=22)**	**Test statistic**	**P-value**	**I (n=22)**	**C (n=22)**	**Test statistic**	**P-value**
Overall performance mean rating	2.9	3.2	$\bar{I} - \bar{C}$ -0.3	0.227†	2.9	3.2	$\bar{I} - \bar{C}$ -0.3	0.229†
Overall performance rating ≥ 3	8.3%	18.2%	Fisher's exact test	0.405	9.1%	18.2%	Fisher's exact test	0.664
Medical management mean rating	2.7	3.3	$\bar{I} - \bar{C}$ -0.6	0.048†	2.6	3.3	$\bar{I} - \bar{C}$ -0.7	0.032†
Behaviour mean rating	2.5	2.9	$\bar{I} - \bar{C}$ -0.4	0.133†	2.5	2.9	$\bar{I} - \bar{C}$ -0.4	0.146†
Mean checklist % score (SD)	43.5 (16.5)	50.5 (16.9)	T -1.422	0.162	43.7 (16.7)	50.5 (16.9)	t -1.346	0.186
Successful (pass mark)	12.5%	13.6%	Fisher's exact test	1.000	13.6 %	13.6 %	Fisher's exact test	1.000
Mean time to O_2 (SD) (sec)	53.3 (22.0)	59.1 (46.3)	Mann-Whit. U 252.5	0.800	54.6 (22.5)	59.1 (46.3)	Mann-Whit. U 227.0	0.725
Mean time to I.V. (SD) (sec)	167.3 (58.5)	168.0 (55.1)	Mann-Whit. U 256.0	0.860	167.1 (57.4)	168.0 (55.1)	Mann-Whit. U -0.054	0.957

* I = Intervention group, C= control group

** $\bar{I} - \bar{C}$ = difference between means

† P-value obtained using Fisher's randomisation t-test

5.3.5. FINAL ASSESSMENT

Fifty three participants (96%) returned for the final assessment which consisted of managing a MOLTI scenario of hypothermia. Data for 53 participants were recovered. No statistically significant differences were found between groups for any of the outcome measures (Table 5.6). In the subgroup of participants for whom a full set of baseline and final assessment data had been obtained, a statistically significant difference was found between the two groups in the mean time to administering I.V. fluids in favour of the intervention group. However, this was not corroborated by any other result and may be the result of multiple statistical testing.

Table 5.6. Final assessment

	All participants				Participants with both baseline and final assessment data			
	I* (n=26)	C* (n=27)	Test statistic	P-value	I (n=22)	C (n=22)	Test statistic	P-value
Overall performance mean rating	2.7	2.8	$\bar{I} - \bar{C}$ -0.1	0.425†	2.8	2.9	$\bar{I} - \bar{C}$ -0.1	0.359†
Overall performance rating ≥ 3	11.5%	18.5%	Fisher's exact test	0.704	13.6%	22.7%	Fisher's exact test	0.698
Medical management mean rating	2.9	2.8	$\bar{I} - \bar{C}$ 0.1	0.418†	3.0	2.8	$\bar{I} - \bar{C}$ 0.2	0.367†
Behaviour mean rating	2.5	2.9	$\bar{I} - \bar{C}$ -0.4	0.087†	2.6	3.0	$\bar{I} - \bar{C}$ -0.4	0.108†
Mean checklist % score (SD)	51.0 (12.9)	50.1 (12.4)	Mann-Whit. U 337.5	0.810	52.6 (13.0)	51.1 (13.0)	Mann-Whit. U 225.0	0.689
Success (pass mark)	23.1%	25.9%	X^2 0.058	0.810	27.3%	27.3%	X^2 0.000	1.000
Mean time to O_2 (SD) (sec)	46.1 (36.5)	40.0‡ (19.2)	T 0.751	0.456	45.3 (39.3)	41.2‡ (18.9)	Mann-Whit. U 222.0	0.827
Mean time to I.V. (SD) (sec)	91.0 (29.7)	104.0 (36.6)	T -1.422	0.161	86.3 (27.6)	106.9 (36.1)	Mann-Whit. U -2.124	0.040

* I = Intervention group, C= control group

** $\bar{I} - \bar{C}$ = difference between means

† P-value obtained using Fisher's randomisation t-test

‡One missing value

5.3.6. FINAL ASSESSMENT SUBGROUP ANALYSES

It was observed that a greater proportion of final assessment videos were lost from the intervention group. The final assessment data were divided into two subgroups based on the type of data that were available to the assessors in each case and the analyses of the checklist and global rating tool data were repeated. No statistically significant differences were found between groups for any of the corresponding outcome measures (Table 5.7).

Table 5.7. Final assessment subgroup analyses on different types of data

	SimMan log data plus video				**SimMan log data only**			
	I* (n=9)	C* (n=15)	Test statistic	P-value	I (n=17)	C (n=12)	Test statistic	P-value
Overall performance mean rating	3.0	3.1	$\bar{I} - \bar{C}$ -0.1	0.500†	2.6	2.4	$\bar{I} - \bar{C}$ 0.2	0.342†
Overall performance rating ≥ 3	33.3 %	33.3 %	X^2 0.000	1.000	0.0%	0.0%	X^2 -	-
Medical management mean rating	3.4	3.1	$\bar{I} - \bar{C}$ 0.3	0.247†	2.6	2.5	$\bar{I} - \bar{C}$ 0.1	0.416†
Behaviour mean rating	2.7	3.3	$\bar{I} - \bar{C}$ -0.6	0.122†	2.5	2.5	$\bar{I} - \bar{C}$ 0.0	0.500†
Mean checklist % score (SD)	62.1 (13.5)	56.1 (12.7)	T 1.099	0.284	45.2 (8.0)	42.7 (7.0)	t 0.862	0.396
Success (pass mark)	33.3 %	33.3 %	X^2 0.000	1.000	17.6%	16.7%	Fisher's exact test	1.000

* I = Intervention group, C= control group

** $\bar{I} - \bar{C}$ = difference between means

† P-value obtained using Fisher's randomisation t-test

5.3.7. AGREEMENT BETWEEN TWO INDEPENDENT ASSESSORS

There was moderate agreement between the two independent assessors in terms of awarding global ratings of <3, 3 or >3 for the final assessment scenario (Table 5.8). Their pass/fail agreement for the same scenario was good. The correlation between the two assessors' global ratings was large for the overall performance and behaviour ratings and moderate for the medical management rating. The correlation between the two assessors' checklist scores was 0.7 (Table 5.8, Fig. 5.5). It is worth noting that the second assessor consistently assigned lower checklist scores than the first assessor as seen in Fig. 5.5.

Table 5.8. Agreement between two assessors - Final assessment

Outcome measure	Kappa	r_s
Overall performance global rating	0.540 (N=49) p-value <0.001	0.584 p-value <0.001
Medical management global rating	0.449 (N=49) p-value <0.001	0.470 p-value =0.001
Behaviour global rating	0.466 (N=48) p-value <0.001	0.536 p-value <0.001
Checklist scores	- (N=52)	0.694 p-value <0.001
Success (Pass/Fail)	0.745 (N=49) p-value <0.001	-

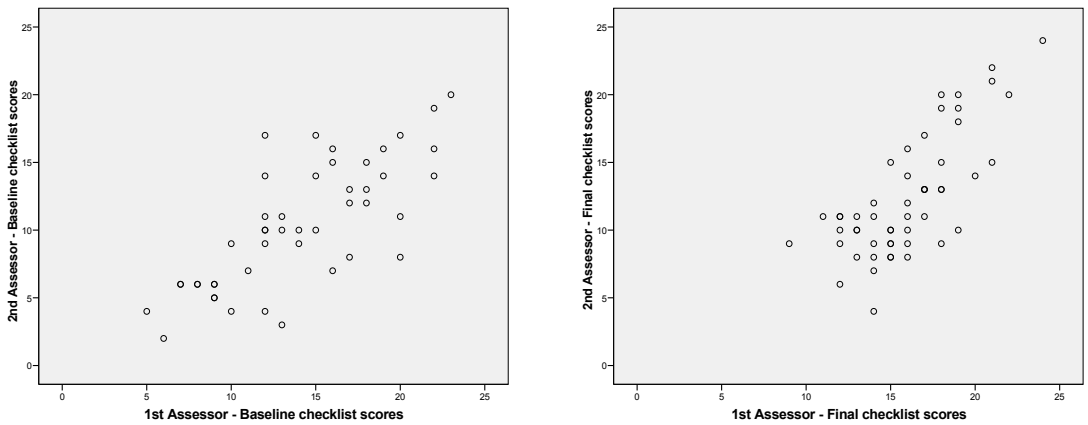

Figure 5.5. Scatter-plots of the two independent assessors' checklist scores

5.3.8. PARTICIPANT SELF-ASSESSMENT

No statistically significant differences were found between the two groups in their five-point Likert-type self-ratings of overall performance at baseline. The difference in the self-ratings of overall performance remained non significant at the final assessment (Table 5.9). No statistically significant differences were found between the two groups' perceptions of their knowledge regarding the assessment scenarios (Table 5.9).

Table 5.9. Participants' ratings of overall performance and scenario knowledge

Self- rated item	Assessment	Intervention mean rating	Control mean rating	$\bar{I} - \bar{C}_*$	P-value
Overall performance	Baseline	1.9 (n=28)	1.9 (n=27)	0.0	0.494
	Follow-up	2.6 (n=26)	2.4 (n=27)	0.2	0.234
Knowledge	Baseline	2.6 (n=28)	2.6 (n=27)	0.0	0.500
	Follow-up	2.9 (n=26)	2.8 (n=27)	0.1	0.382

There was a significant 'before and after' difference, from baseline to follow-up after MOLTI training for all the participants in their responses with regards to confidence in starting the right basic management if called to see a deteriorating patient. For the corresponding question the mean Likert-type score rose from 2.53 (N=41) at the baseline questionnaire to 3.66 (N=41) at the feedback questionnaire following MOLTI training ($Z = -4.815$, $p<0.001$).

5.3.9. COMPARISONS OF PARTICIPANT AND RATER ASSESSMENTS

When comparing the participants' self-ratings of overall performance to their two raters', more participants underrated their performance during the baseline assessment than during the final assessment. This trend appeared to be stronger in the intervention group (Fig. 5.6). During the final assessment the overall performance score agreement between the participants and each rater was not statistically significant ($kappa_1 = 0.080$, $p_1=0.363$ and $kappa_2 = 0.123$,

p_2=0.085). The correlations between participants and each rater were small and not statistically significant (r_{s1}=0.093, p_1=0.508 and r_{s2} = 0.201, p_2=0.167).

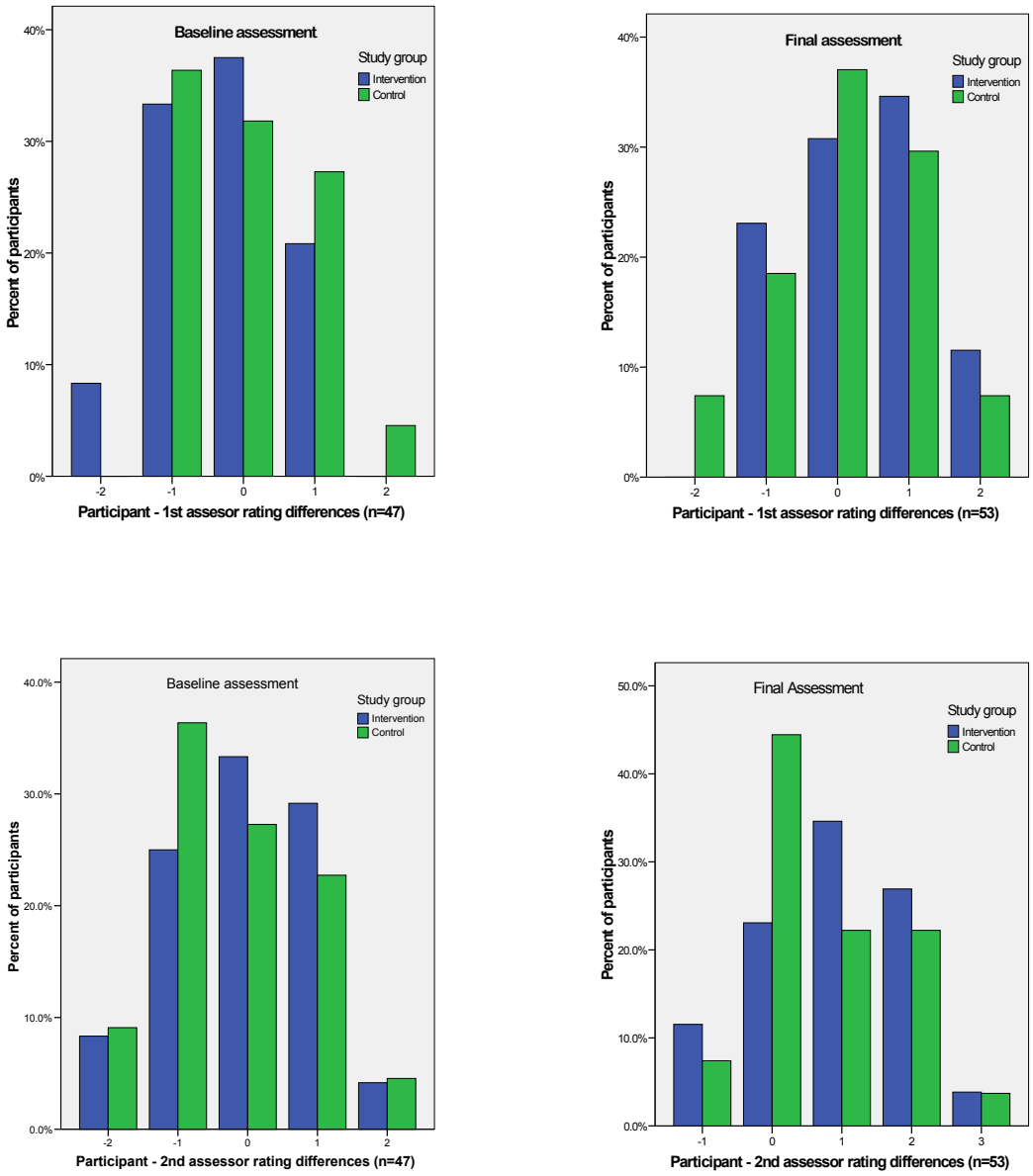

Figure 5.6. Differences between participant self-ratings and assessors ratings for overall performance (difference = participant rating – instructor rating)

5.3.10. PARTICIPANT FEEDBACK

53 participants (96%) returned their feedback questionnaires. Ninety-eight percent of the respondents 'agreed' or 'strongly agreed' that their MOLTI training had been useful and 2% 'disagreed'. Their responses in the Likert-type feedback questions regarding the MOLTI course are presented in Table 5.10.

Table 5.10. Percentage of participants' responses to given statements following the MOLTI course (n=53)

Statement	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I found the orientation to the simulator adequate	0	7	15	65	13
The information provided by the instructor during the baseline scenario was clear	0	2	4	67	27
I was able to manage the baseline scenario based on reading/general knowledge without prior experience	11	53	24	7	5
The information provided by the instructor during the final assessment scenario was clear	0	0	4	57	39
What I learned from the scenarios was more than the knowledge I brought to it	0	4	21	47	28
The simulator enhanced learning more than reading would	0	0	2	34	64
I felt I did things I would never have a chance to practice otherwise	0	9	9	57	25
The debriefing provided logically organised feedback and clarified important issues of the scenario	0	0	2	57	41
The debriefing provided enhanced my stock of knowledge	0	2	11	45	42
I expect that the knowledge gained from the scenarios will be helpful to me in practice	0	0	2	49	49
If I was called to see a deteriorating patient, I am confident I could start the right basic management	0	9	28	51	11
If I was called to a deteriorating patient I am worried I would do the wrong thing	2	32	25	32	9
I felt comfortable with the simulator environment	0	17	13	51	19
I felt that the simulation environment and scenarios prompted realistic responses from me	2	11	23	45	19
I found it easy to treat the mannequin as a simulated human	6	34	19	32	9

Table 5.10. Percentage of participants' responses to given statements following the MOLTI course (continued)

Statement	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
The presence of the scenario director detracted from the realism of the simulator experience	13	55	15	15	2
The video camera interfered with the simulator experience	38	53	6	2	2
Overall, I was satisfied with the use of the human patient simulator for training*	0	2	4	60	34
I found the MOLTI course useful	0	2	0	34	64
I enjoyed the MOLTI course	0	0	7	36	57

*N=52

Fifty-one of the respondents (96%) opted to include some further qualitative feedback.

5.3.10.1. FEATURES OF THE MOLTI COURSE PERCEIVED AS POSITIVE

With regards to positive MOLTI course features, the main themes that were mentioned by the participants in the qualitative feedback are presented in Table 5.11.

Table 5.11. Features of the MOLTI course that were perceived as positive (N=51)

Theme	n (%)	Example quotes
The opportunity to practice	20 (39.2)	*'Good practice as reading it from a book is not like real life.'* *'Being in a simulated environment was the nearest thing to managing a deteriorating patient – it was helpful to practice how I would respond.'*
The MOLTI course promoted learning	16 (31.4)	*'It helped me consolidated my knowledge and highlighted the importance of basic resuscitation.'* *'Has really helped me learn about how to manage or approach sick people.'*
The content of the MOLTI scenarios	15 (29.4)	*'Revisiting and emphasis on the basics. Good simulation of real life scenarios – good examples of scenarios also.'* *'The scenarios are common situations that we will encounter in the future.'*
The use of simulation	12 (23.5)	*'The use of a simulator made the course feel very realistic.'* *'The different range of scenarios and in screen with monitoring helped.'*
The provision of feedback	11 (21.6)	*'Good to go through the scenarios after to show you where you went wrong. Learnt a lot.'* *'Feedback – something we get very little of.'*

Table 5.11. Features of the MOLTI course that were perceived as positive (continued)

Theme	n (%)	Example quotes
The realism of the course	10 (18.6)	*'Being faced with a realistic scenario, whereby I have to initiate the management of critical patient.'* *'Being in a simulated environment was the nearest thing to managing a deteriorating patient.'*
Systematic approach to managing life-threatening illness	9 (17.6)	*'Highlighted things I knew already but wasn't able to apply – helped to think of it in a systematic way.'* *'Helpful to go through steps in management of deteriorating patient – gives a framework to fall back on when confronted by real life situation.'*
The teaching	9 (17.6)	*'The instruction was clear, relevant and practical for our stage of training.'* *'Teaching on scenarios. Teaching was aimed at the appropriate level. Best teaching all year.'*
Being 'put on the spot'	8 (15.7)	*'Nothing like being put on the spot to help sear situations in your memory!'* *'Being put in high pressure situations helped me to think on my feet, won't always be someone around to help.'*
Made participants think	5 (9.8)	*'Brilliant simulator mannequin – learnt how to manage acutely ill patients with good scenarios that were realistic, so it actually made us think.'* *'Makes you think independently and quickly.'*
Seeing the effects of interventions	4 (7.8)	*'Good to see effects of our 'treatment' decisions.'* *'The facility to see straight away the effect of my management decisions.'*
One-to-one approach	4 (7.8)	*'The one-to-one element makes you learn more. Adds pressure to the situation as supposed to just chatting through what you would do.'*
The MOLTI course increased participant confidence	4 (7.8)	*'Helped give me confidence as I realised I knew a bit more than I thought.'* *'Being put on the spot with real time obs and updates on patients status/wellbeing. Tutorial very useful. Both acted to improve my confidence of managing acute scenarios.'*
Chance to assess their knowledge	3 (5.9)	*'Pressure to do things on the spot and test myself rather than just relying on knowledge from books and observing in hospital.'* *'Having a chance to see what you know and what you don't.'*
The course was non-intimidating	2 (3.9)	*'The atmosphere was very friendly and non-intimidating which helped put you at ease in a false situation.'* *'Everyone [was] friendly and made me comfortable to scenario.'*
A unique practice opportunity	2 (3.9)	*'Really useful to have experience real-time & emergency situation – can't get elsewhere.'* *'Exposure to situations previously unexperienced.'*

5.3.10.2. LIMITATIONS OF THE MOLTI COURSE AND SUGGESTIONS FOR FURTHER IMPROVEMENT

Qualitative feedback on MOLTI course limitations and suggestions for further improvement are presented in Table 5.12.

Table 5.12. MOLTI course limitations and suggestions for further improvement (N=51)

Theme	N (%)	Example quotes
Difficult to treat the simulation as real	5 (9.8)	*'I felt the fact that the model is not real may have put me off, but I still feel the overall benefits and learning is invaluable.'* *'It's difficult to grasp at the beginning that you should treat it like a human.'*
Extend the MOLTI sessions	19(37.3)	*'Longer teaching session, more practice with different scenarios.'* *'Perhaps more personalised feedback and advice about what individually to improve on.'*
Further familiarisation with the simulator	9 (17.6)	*'Maybe see the SimMan in action in a scenario so that [we] fully understand what we can do.'* *'Explain how far to go in the scenario in terms of tests. It was difficult to know how seriously to take the simulation.'*
Integrate MOLTI in the MBChB curriculum	8 (15.7)	*'Very worthwhile, I think it should be part of the standard 5th year curriculum or even part of the assessment. Would definitely recommend them.'* *'This would be a great teaching tool to use for all 4th and 5th year students – especially in ARICM and Surgery blocks'*
Provide opportunity to practice on their own time	3 (5.9)	*'Perhaps additional access to the scenarios on own to practice afterwards in own time.'*
Provide handout	3 (5.9)	*'Give handouts on the steps on how to deal with common scenarios.'*

Individual suggestions included encouraging participants to talk through what they are doing so that thought processes are clear, including psychomotor skills practice during the scenarios, teaching further management following the primary resuscitation phase and providing the opportunity to view themselves on video as an additional component to feedback.

5.4. DISCUSSION

Main findings

The purpose of this study was to investigate the efficacy of the MOLTI course using the SimMan. In the final assessment, the participants who received MOLTI training achieved similar global ratings, checklist scores and times to specific interventions to those who received no such training. No statistically significant differences were found between the two groups in their self-ratings of overall performance. Agreement between participant ratings and their instructors' ratings was poor.

However, the majority of participants felt that the MOLTI course was useful. Furthermore, there seems to be a significant 'before and after' difference in the participants' confidence, from baseline to follow-up after MOLTI training. This is further reflected on the differences between participant and instructor ratings; more participants in the intervention group overrated their performance following MOLTI training compared to baseline. According to their qualitative feedback, participants valued the opportunity to practice and the MOLTI scenarios, and felt that the MOLTI session promoted learning. Participants suggested that the MOLTI sessions be extended to include more opportunities to practice, longer teaching, more feedback, more scenarios and a more detailed familiarisation session with the simulator.

Limitations

Due to a technical breakdown of the data storage equipment a substantial part of the videos were lost. Thus, some cases were rated using both a video recording and the corresponding SimMan log, while others were rated using the SimMan log only. This appears to have had an impact on the assessors' evaluation. The assessors rated five participants' hypothermia

scenarios twice; once using only the SimMan log sheet and once viewing the corresponding video. Their checklist scores and global ratings increased after viewing the corresponding videos which may imply that lack of a video recording biased scores downwards. More videos of the final assessment were available for the control group than the intervention group (ratio 15:9), which could have masked the effectiveness of the MOLTI course. In an effort to address this limitation the affected data were assigned to two separate subgroups and all analyses were repeated; one subgroup included cases with a video recording and SimMan log while the other subgroup included cases with a SimMan log only. Again, in both subgroups the participants who received MOLTI training had similar global ratings and checklist scores to those of the participants who had received no such training.

Furthermore, there was no attempt to compare pre-intervention to post-intervention scores given the differences between the two scenarios used. These assessment scenarios were subjectively judged to be equally difficult by a consultant with expertise in simulation, however creating scenarios of equal complexity is challenging. In the final assessment, the SimMan may not have been successful in portraying realistically the selected scenario. This scenario depicted an elderly male who had been brought to hospital after having been found unconscious in his bath and was suffering from hypothermia. The fact that participants were faced with an inanimate mannequin that lacked external signs e.g. skin colour may have hindered some of them from assessing the 'disability' component of the scenario. Simulator fidelity may have been inadequate for such a scenario.

Consideration of possible mechanisms and explanation

A number of non-exclusive reasons might explain the null results of this study. Firstly, the unilateral loss of final assessment videos due to technical breakdown may have led to the control group receiving higher ratings by the assessor than the intervention group. This effect may have concealed any positive effects of the MOLTI course training. However, the subgroup analyses of the data did not reveal a consistent trend towards either direction. Furthermore, the times to specific interventions, which were not affected by the loss of videos, were not significantly different between the two groups.

Secondly, each assessment could be deemed as a learning experience in its own right. Thus, learning resulting from the baseline assessment simulation experience may constitute another explanation of why the intervention and the control group achieved similar ratings. However, the absence of feedback following the baseline assessment in the case of the control group makes this explanation less likely.

Thirdly, the MOLTI course was not run as an incorporated part of the curriculum during this study, thus it may have been out of context. Within the MBChB curriculum the MOLTI course is delivered following ILS training. However, due to scheduling incompatibilities, it was not possible to implement the study immediately following the participants' ILS training. All students had had basic life support training but their knowledge base may not have been adequate to build on with the MOLTI course. In addition, the half hour duration of the course and its 'one-off' nature may not have been enough to consolidate its teaching principles.

Finally, the simulator may have subjectively improved the training experience of the participants and enhanced their confidence without objectively altering test performance. Our results in this case would be in agreement with the RCT by Cherry et al[131]. Previous studies have shown that simulation-based learning is highly valued by medical students[125]. A positive influence on participants' perceived confidence following simulation training has been previously reported[124]. Another study found a significant correlation between the number of times a skill was performed and the student level of confidence but no correlation between level of confidence and clinical grades[132]. It may be that students learned how to deal with the crises to which they were exposed, but that they were not able to adapt their new skills to a different crisis. It could be that practice on the simulator leads to increased confidence without affecting performance. However, increased confidence might still be of value in clinical practice even if outcomes are not affected; confident students might approach their clinical practice more enthusiastically and pursue practice opportunities more actively.

Summary of clinical and research implications

The majority of participants rated the MOLTI course and their simulation experience favourably and thought it should be integrated in their final year curriculum. Thus, a next step might be to refine the existing MOLTI course by addressing student comments regarding extended teaching, feedback and simulator familiarisation. The creation of a course handbook may also be warranted to help students revise and reflect on their MOLTI training experience. Future studies should then seek to further explore the efficacy of the MOLTI course using larger sample sizes. *In vivo* evaluations are also needed to investigate the consequences of such training to patient outcomes, although it is appreciated that this is difficult to do in practice.

Individual instructors should seek to include computer-controlled mannequins in MOLTI teaching if one is available in their Trust. However, further research is needed to compare SimMan practice to alternative methods delivering the same course material such as e-learning. Future studies should also investigate the costs of providing computer-controlled mannequin training, including opportunity costs, and explore whether this training makes the best use of instructor, student and simulator time.

CHAPTER 6

DISCUSSION: IMPLICATIONS FOR PRACTICE AND RESEARCH

6.1. CHAPTER OVERVIEW

The present chapter aims to provide an overall synthesis of the theory and research that was reviewed and undertaken during this thesis. A framework for helping educators to determine whether simulation is an appropriate medium within a well-designed training programme is proposed. This framework is then used to synthesise the research gaps addressed, the limitations of the studies undertaken and the research gaps that it may be useful to address in the future. The chapter concludes with implications for practice and research.

6.2. SIMULATION AS A TRAINING MEDIUM

The educational theories identified in Chapter 2 and the research reviewed in Chapter 3 suggest that simulation training can be effective both in its own right and in comparison to other methods of training. However, it is clear that simply including simulation in a training programme will not guarantee its effectiveness. In a simulation course *simulation* is only one of the elements- the medium- that needs to be taken into consideration when trying to accomplish effective training. Other elements include the configuration of the training programme (e.g. small-group or self-practice), mode of delivery (e.g. trainer-led or student-led) and the instructional methods used (e.g. self-assessment, group discussion, feedback)[27, 133]. It is easy to see how the effectiveness of a training course based on a well-designed simulation could be undermined by poor quality instruction, for example. Furthermore,

interactions between the *simulation* element and the other elements of the training course may influence its resulting effectiveness. For example, in the case of the group size element interacting with the medium element, simulation is only feasible if there are enough simulators to go round.

6.2.1. A DECISION FRAMEWORK FOR TRAINING IN HEALTHCARE

While this thesis has considered simulation training as a possible training medium, the key question for those designing a training programme is not 'should I use simulation?' but 'how can I maximise the effectiveness of this training?'. In order to establish whether the simulation medium is in fact an appropriate choice for the skill to be taught, a series of questions might be helpful for educators during the design and planning phase of the training programme. Simulation therefore needs to be considered within this wider decision framework, an example of which is shown in Fig. 6.1. A set of nationally recognised guidelines for the use of the different types of simulation training would subsequently help educators understand each one's potential and ensure good practice: answering question 5.1 in the framework.

At the outset, educators need to be aware of the educational theories and course design principles presented in Chapter 2 and engage in critical reflection on educational research evidence, theory and their professional practice. The Postgraduate Certificate and Associate Programmes in Learning and Teaching in Higher Education, which are offered to educators by most UK universities and are accredited by the UK Academy of Higher Education, teach such an approach (including course design) and might be valuable in promoting good educational practices [134].

1. What are the intended learning outcomes of the training and at what level of the Miller hierarchy do you want to see results?
2. What standard are trainees expected to reach?
3. Do trainees have the underpinning knowledge and skills required?
4. What resources are available and what are the feasibility constraints?
5. What medium should be used? Should this be simulation?
    - 5.1. If so, what type of simulation?
    - 5.2. How much simulation is needed?

Figure 6.1. Decision framework for training in healthcare

6.2.1.1. LEARNING OUTCOMES

At the first level of the decision framework (Fig. 6.1), educators need to identify and clearly define the learning outcomes of the training programme at hand[51], as noted by Issenberg et al as an important prerequisite for effective learning[16] and discussed in section 2.5. Defining the detailed steps of the skill being taught is a prerequisite to training[56]. In the RCT in Chapter 4, participants completing the training intervention were expected to be able to demonstrate ('shows how' in Miller's pyramid) an LMA placement on a part-task trainer. The evaluation of the intervention considered participants' ability at the 'does' level in Miller's pyramid by assessing performance in clinical practice. In the RCT in Chapter 5 participants completing the MOLTI course were expected to be able to understand ('knows how' in Miller's pyramid) the importance of reversing adverse physiological trends quickly, practice ('shows how' in Miller's pyramid) managing life-threatening illness on an integrated simulator and transfer these skills to different scenarios.

6.2.1.2. TRAINING STANDARDS

Another consideration regarding simulation-based learning is the development of clinical benchmarks as a standards-setting mechanism to establish skill competency and the clinical relevance of simulation[135]. These benchmarks relate to meeting the learning outcomes for the training and may or may not correlate with subsequent 'in practice' outcomes. An example approach to standard-setting is determining the minimum level of performance considered acceptable by a panel of one or more experts[135]. In the RCTs described in chapters 4 and 5, it was not possible to set such benchmarks due to practical constraints. In the first RCT, participants were asked to practice until they had demonstrated a correct LMA placement based on their instructors' judgement. The second RCT included no standards-setting mechanism to establish skill competency during the training intervention. Instead, participants practiced on a set number of MOLTI scenarios, receiving immediate feedback on their performance.

6.2.1.3. BACKGROUND KNOWLEDGE

High-fidelity simulation requires that trainees already possess the underpinning knowledge and skills that will be required during the simulation[27], something that was not always the case with the MOLTI participants of the RCT in Chapter 5. Both the two raters and the trainer felt that the participants' knowledge-base was lacking for their stage in the MBChB curriculum. It may have been the case that the MOLTI scenarios were too advanced for undergraduate medical students. A possible solution would be to revise the scenarios in line with the MBChB curriculum or to undertake the MOLTI course at a later stage in the curriculum. The need to ensure that simulation is used at the appropriate stage in the curriculum is noted by Alinier[27]. Kneebone highlights the importance of ensuring that

simulation training is related to learners' everyday work in order to allow its incorporation into clinical practice[42]. At present, the MOLTI course is offered as an 'add on' to the Year 5 curriculum in Surgery and hence integration of the MOLTI simulation element into the curriculum may help ensure that students have an adequate knowledge base to build on in order to achieve transfer of skills from one scenario to another. In the case of the RCT in Chapter 4 the majority of participants had already received basic LMA training (and therefore built on their existing knowledge) and the training intervention was related to their clinical placement at the time (enabling skills to be incorporated into clinical practice).

6.2.1.4. RESOURCES AND FEASIBILITY

The feasibility and cost of including all of the features identified by Issenberg et al as leading to effective learning[16] is an important consideration when designing an educational programme. The optimal combination of these features is likely to vary across training programmes and a reflective cycle of programme design may be required to identify the most effective training that can be delivered within the feasibility and resource constraints of the programme. For example, it would not be possible for an entire cohort of MBChB students to undertake the MOLTI course at the end of their final year. While this would mean they would all have the 'full' background knowledge and experience to help them maximise their learning gains from the course, the resources required would be prohibitive. Each SimMan costs around £27000[29] and it would not be cost effective to buy sufficient simulators for this intensive training period if they would not be used for much of the rest of the year.

6.2.1.5. CHOOSING THE TRAINING MEDIUM

It is important to consider when simulation may be particularly effective as a training medium within a well-designed training programme. A useful approach might be to consider the desired learning outcomes against Miller's pyramid and then do the same for the available training mediums (Table 6.1). This would help trainers to identify the extent to which each of the available mediums can produce the desired learning outcomes and hence match the training programme to its ideal training medium. Simulation should therefore be chosen if it is best placed to enable trainees to meet the learning outcomes of the training programme[136]. With regards to the RCTs reported in this thesis, the simulation medium could be considered a sensible choice as both training interventions included a 'shows how' practical skills learning outcome (see section 6.2.1.1).

Table 6.1. Example of mapping between *Tomorrow's Doctors* [19] learning outcomes, Miller's pyramid and training mediums

Example learning outcomes from *Tomorrow's Doctors*	Miller's pyramid	Example choices of medium
Provide explanation, advice, reassurance and support during patient consultation	Does	Bedside practice Simulator
Take and record a patient's medical history	Shows how	Bedside practice Simulator
Explain the fundamental principles underlying investigations for common clinical cases	Knows how/ Understands	Lecture Small-group tutorial Books/ Journals
Explain normal human structure and functions	Knows	Lecture E-learning tutorial Books/ Journals

The degree of difficulty and the type of the task at hand (technical, cognitive or social/interactive) may also be important factors when determining whether simulation is the most effective medium. Based on the review of reviews in Chapter 3, VR simulation appears to be effective both *in vitro* and *in vivo* for training in the 'complicated' laparoscopic skills. However, laparoscopy and other minimally invasive tasks can be easier to simulate than open surgery due to the limited visual and haptic feedback involved *in vivo*[90]. The findings of the review of reviews prompted the hypothesis that the need for simulation training may depend on the degree of difficulty of the task, with simulation training being more effective as the task difficulty increases. This hypothesis appears to be supported by the RCT described in Chapter 4, where an additional session of mannequin training did not lead to better real life performance with regards to LMA insertion, which is considered a relatively simple technical skill to acquire.

The generation of a scale of task 'difficulty' would be useful in order to study the relationship between difficulty of task and effectiveness of simulation training. The dimensions for such a scale could include component complexity (e.g., number of steps in the procedure), coordinative complexity (requirements of timing and sequence), dynamic complexity (changes in the task over time), cognitive complexity (knowledge recall and decision-making), team interactions, affective issues, external (system) complexity and time to completion[56]. An alternative approach to assessing task difficulty could utilise Kneebone's traffic light model which defines three zones of risk for learning any procedure or intervention (low, moderate and high)[42].

6.2.1.6. CHOOSING THE TYPE OF SIMULATION

When choosing which type of simulator to use, the nature of the task to be learned should be considered in combination with the potential of different simulators. For example, computer-based simulators may be better for practicing an invasive technical procedure, while standardised patients may be better for practicing management of mental illness. It has been argued that fitness for educational purpose should take precedence over simulator complexity in choosing between types of simulator[42]. For example, the simple mannequins used in the RCT in Chapter 4 were suitable for teaching the process of LMA insertion, but were considered to mimic reality poorly and hence were not suitable for teaching insertion skills.

In order to establish which type of simulator is warranted for each level of skill acquisition in Miller's pyramid, Alinier has proposed a framework for acquisition of experience and skills through simulation training[27]. This framework (as shown in section 1.3.2) could serve as a useful guide for educators trying to choose which type of simulator to use. In his framework, Alinier implies that the need for simulation fidelity and complexity increase as one progresses from acquiring knowledge to engaging in action[27]. This framework therefore complements the hypothesis that the need for any type of simulation increases as the difficulty of the task increases and implies that 'difficult' skills to be acquired at the 'does' level of the Miller's pyramid are the most likely to require high fidelity simulators. However, as Norman interestingly points out, thought should also be given to the *kind* of practice required for the specific skill being learned[137]. Norman highlights three examples that could challenge Alinier's framework with regards to mastery; a) mastering suturing requires many trials on a low-fidelity simulator, b) mastering auscultation of different heart sounds could be accomplished nearly as well with a CD-ROM as with a high-fidelity simulator, and c) experts

tuning up their reasoning skills could achieve deliberate practice over coffee without requiring a physical environment, simulated or otherwise, like chess masters do not require a board to play on.

Overall, educators considering including simulation among the elements of a training course should ponder whether the simulation element is justified. For example, the simulation training offered in the Chapter 5 RCT could be justified as the vast majority of the MOLTI participants felt that the simulation course offered was in fact useful and they felt more confident following this training. This course may therefore still have the potential to lead to a change of attitudes, which is part of the Kirkpatrick training criteria presented in section 3.3.3. An *in vivo* follow-up of this study would be the ideal way to test whether the MOLTI training affected the trainees' clinical management while facing similar problems as Foundation Year 1 doctors, but this was precluded due to practical considerations.

6.2.1.7. CHOOSING THE AMOUNT OF SIMULATION

It has been recommended that the 'dose' of practice necessary should be determined by the trainees' need rather than the instructor's availability and views of the trainees' need[47]. The review of reviews reported in Chapter 3 highlighted that there is limited evidence regarding the optimal duration of simulation as part of a training programme. However, the general trend was in the form of a positive dose-response relationship. The RCT presented in Chapter 4 compared the effectiveness *in vivo* of two LMA placement simulation courses of different duration. A previous review had identified that there is little literature on actual methods of training for this technique[114]. In contrast to the evidence reported in Chapter 3, this RCT

found no additional benefit from the additional training provided for the intervention group, possibly due to the relatively simple nature of the task.

6.2.2. ACHIEVING LEARNING TRANSFER

The RCT reported in Chapter 5 found that simulation training did not result in better *in vitro* performance in terms of the cognitive skills required to manage life threatening illness. The ability to use 'a previously learned concept to solve a new, apparently different problem' or *transfer* is a complex topic in medical education, which is still under investigation[138]. Students may have learned how to deal with the individual crises to which they were exposed during the intervention training scenarios, but may have not been able to adapt this knowledge to the different scenario used during the final assessment. Hence, more research is required regarding how simulation training can assist in the transfer of skills across tasks and between the training environment and clinical practice environment.

6.2.3. CONCLUSION

This thesis has added to the literature in the field of medical education a review of reviews of the evidence regarding the effectiveness of simulation training in medicine and surgery; and two RCTs evaluating different simulation training courses. The review of reviews highlighted that simulation training can be effective, but there was little consistent evidence across tasks or types of simulator. The two RCTs reported nil results, reinforcing that simulation alone is insufficient to ensure effectiveness. These results highlight the importance of recognising when simulation training is appropriate, how simulation interacts with other elements of a training programme and how the simulation can be made maximally effective.

Before providing some implications for both research and practice, it is necessary to bear in mind that simulation training is an ever developing field. Further technological advances are needed in order to increase simulator fidelity, ease of use and accessibility to trainees. At the present stage, even in laparoscopic surgery, the advances in technology have not reached the stage where simulation training could replace real-life training altogether. Simulation cannot replace the situational context and complex interactions learned trough interaction with real patients[15]. Such developments make continuous research important to ensure training continues to be effective and to identify avenues for improvement.

6.3. IMPLICATIONS FOR PRACTICE

- The design of simulation training should be based on sound educational theories as the principles of effective learning apply in this case as they do for other training media.
- The instructional design, the delivery methods and the features of simulation training should be carefully selected based on the type and degree of difficulty of the task at hand.
- Trainees should be made explicitly aware of the anatomical, texture and clinical variations between the simulator used and real life.
- Trainees should be given clinical benchmarks to achieve as a standards-setting mechanism to establish procedural skill competency during formative assessment.
- For relatively simple and non-invasive technical skills such as LMA insertion, it may be more cost-effective in terms of training outcomes to train under supervision 'at the bedside' rather than use simulation.
- A set of nationally recognised guidelines for the use of the different types of simulation training would help ensure good practice.

6.4. IMPLICATIONS FOR RESEARCH

- Further research is needed to test *in vivo* the hypothesis that the effectiveness and hence the need for simulation training may depend on the type and degree of the difficulty of the procedure in question.
- The generation of a scale of task 'difficulty' is needed in order to study the relationship between difficulty of task trained and the effectiveness of simulation training.
- Future studies should seek to evaluate whether simulation training leads to skills retention over time.
- Further research is needed to evaluate the transfer of skills learned through simulation to real life situations.
- Studies are needed to evaluate the impact of simulation training on patient outcomes such as patient care and medical error.
- Further studies are also needed to compare the effectiveness of simulators of varying fidelity in training for a specific task with a view to determine the most cost-effective simulators in terms of training outcomes. Fidelity may be a determinant of cost-effectiveness with higher fidelity leading to extra cost, while time is another key issue e.g. the opportunity cost of trainees' time. A full cost-effectiveness analysis would be warranted.
- Studies investigating simulation effectiveness should specifically address the validity and reliability of their assessments.
- Education studies should include contemporaneous controls where possible.

APPENDICES

APPENDIX 3.1. CRITICAL APPRAISAL RESULTS – EXCLUDED REVIEWS

Table 3.2. Critical appraisal of excluded reviews (√ = Yes, X = Not reported, - = the review format precludes judgment)

	Stage one					Stage two			
Author and date	Specifies clear aim or research question	Identifies appropriate range of source databases	Undertakes additional search strategies*	Specifies search terms	Specifies inclusion criteria	Rigour of individual studies assessed	Individual studies' findings presented clearly and consistently	Individual studies' findings analysed clearly and consistently	Conclusions presented relate to individual studies' findings
Aggarwal et al 2004[91]	X	X	√	√	X	√(in part)**	X	X	–
Eppich et al 2006[139]	X	X	X	X	X	√(in part)**	X	X	–
Fitzerald et al 2008[140]	X	X	X	X	X	X	X	X	–
Goldmann et al 2005[141]	X	X	X	X	X	X	X	X	–
Gould et al 2006[142]	√	X	X	X	X	√(in part)**	X	X	–
Hammond & Karthigasu 2006[89]	X	X	X	X	X	X	X	X	–
Hamstra et al 2006[34]	√	X	X	X	X	X	X	X	–

*Additional search strategies involve follow-up of references/journals, consultation with experts in the field and searching for grey literature

**i.e. rigour was addressed only for some of the studies described or for a specific aspect of the studies

Table 3.2. Critical appraisal of excluded reviews (√ = Yes, X = Not reported, - = the review format precludes judgment) (continued)

	Stage one					Stage two			
Author and date	Specifies clear aim or research question	Identifies appropriate range of source databases	Undertakes additional search strategies*	Specifies search terms	Specifies inclusion criteria	Rigour of individual studies assessed	Individual studies' findings presented clearly and consistently	Individual studies' findings analysed clearly and consistently	Conclusions presented relate to individual studies' findings
Hart & Karthigasu 2007[143]	√	X	X	X	X	X	X	X	–
Jha et al 2001[36]	√	X	X	X	X	X	X	X	–
Khalifa et al 2006[94]	X	X	X	X	X	X	X	X	–
Laguna et al 2006[144]	X	X	X	X	X	√(in part)**	√(in part)	X	–
Laguna et al 2002[95]	X	X	X	X	X	√(in part)**	√(in part)	X	–
Lake 2005[92]	√	X	X	X	X	X	X	X	–
Lamb 2007[145]	√	X	X	√	X	X	X	X	–

*Additional search strategies involve follow-up of references/journals, consultation with experts in the field and searching for grey literature

**i.e. rigour was addressed only for some of the studies described or for a specific aspect of the studies

Table 3.2. Critical appraisal of excluded reviews (√ = Yes, X = Not reported, - = the review format precludes judgment) (continued)

	Stage one					Stage two			
Author and date	Specifies clear aim or research question	Identifies appropriate range of source databases	Undertakes additional search strategies*	Specifies search terms	Specifies inclusion criteria	Rigour of individual studies assessed	Individual studies' findings presented clearly and consistently	Individual studies' findings analysed clearly and consistently	Conclusions presented relate to individual studies' findings
McFetrich 2006[37]	X	√	X	√	√	X	X	X	–
Michelson 2006[146]	√	X	X	X	X	X	X	X	–
Rosenbaum et al 2004[147]	√	X	X	X	X	X	√	√	–
Seymour 2008[148]	√	X	X	X	X	X	X	X	–
Stringer et al 2002[114]	X	X	X	X	X	X	X	X	–
Undre & Darzi 2007[149]	√	X	X	X	X	√	X	–	–
Villegas et al 2003[150]	√	X	X	X	X	X	X	X	–
Wong 2004[67]	√	X	X	X	X	X	X	X	–

*Additional search strategies involve follow-up of references/journals, consultation with experts in the field and searching for grey literature
**i.e. rigour was addressed only for some of the studies described or for a specific aspect of the studies

APPENDIX 3.2. REVIEW OF REVIEWS DATA: INDIVIDUAL STUDIES PERTAINING TO SURGERY (TABLES 3.3 – 3.16)

Table 3.3. Summary of the individual studies reviewed pertaining to model surgical training compared to no training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Physical Trainer/model training vs no training –in vivo							
Hamilton et al 2001[77, 83]	RCT	Laparoscopic hernia repair *in vivo* (OR)	Ni=11 Nc=11 Junior surgical residents	Model training (video, CD-ROM, rubber hernia simulator) vs no training	Operative performance	Yes	Performance: 5 out of 8 assessment areas (√)
Pohl et al 2003[83]	RCT	Laparoscopic cholecystectomy *in vivo* (pig)	N=28 Medical students	Synthetic material simulator training (2 simulations) vs repeated synthetic material simulator training (until plateau) vs no training	Laparoscopic cholecystectomy performance	Not stated	Repeated synthetic material simulator training (until plateau): Performance (√)
Youngblood et al 2005[77, 83]	RCT	Laparoscopic performance *in vivo* (pig)	Ni=16 Nc=13 46 medical students	TowerTrainer training vs no training	Laparoscopic performance	Yes	Performance (-): in 1 out of 7 outcomes(√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.4. Summary of the individual studies reviewed pertaining to model surgical training compared to standard training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Inanimate model training vs standard training –in vitro							
Anastakis et al 1999[77, 83]	RCT	Surgical tasks *in vitro* (cadaver)	N = 23	Bench model training vs standard training (independent learning from manual)	Performance of 6 surgical procedures (global rating, checklist scores)	Yes	Performance (√)
Matsumoto et al 2002[77, 83]	RCT	Endourological skills *in vitro* (video-box trainer)	Ni=16 Nc=7 40 medical students	Bench model training vs didactic training (1hour instruction)	Performance scores (global rating, checklist scores, achieving pass rating) Time taken	Yes	Performance (√) Time taken(√)
Inanimate model training vs standard training –in vivo							
Grober et al 2004[83]	RCT	Urological microsurgery *in vivo* (animal)	N=18 Junior surgical residents	Bench model training vs didactic training	Performance at 4 months	Yes	Performance (√)
Velmahos et al 2004[82, 83]	RCT	Central venous catheterisation *in vivo* (patients)	N = 26 Surgical interns	Mannequin lab training vs bedside training	Number of attempts to find the vein Time to completion Knowledge (MCQ) Competence in catheterisation (Checklist)	Yes	Number of attempts (√) Time (-) Knowledge (√) Competence (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.5. Summary of the individual studies reviewed pertaining to model surgical training compared to other types of simulation training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box training vs model training –in vitro							
Matsumoto et al 2002[77, 83]	RCT	Endourological skills *in vitro* (video-box trainer)	N1=17 N2=16 40 medical students	Video-box training vs low fidelity bench model training	Performance scores Time taken	Yes	Performance (-) Time taken (-)
Model training vs cadaver training –in vitro							
Anastakis et al 1999[77, 83]	RCT	Surgical tasks *in vitro* (cadaver)	N = 23	Bench model training vs cadaver training	Performance of 6 surgical procedures	Yes	Performance (-)
VR training vs Physical Trainer training/model training- in vivo							
Youngblood et al 2005[77, 83]	RCT	Laparoscopic performance *in vivo* (pig)	N1=17 N2=16 46 medical students	LapSim VR training vs TowerTrainer training	Laparoscopic performance	Yes	Performance in 3 out of 7 outcomes(√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.6. Summary of the individual studies reviewed pertaining to video-box surgical training compared to no training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box training vs no training -in vitro							
Jordan et al 2001[74, 77, 83]	RCT	Laparoscopic skills *in vitro* (box trainer)	Nz=8 Nu=8 Nc=8 32 medical students	Video-box training (Z-maze) vs video-box training (U-maze) vs no training	Performance (correct incisions)	Not stated	Z-maze training: Performance (√)
Munz et al 2004[77]	RCT	Clip application *in vitro* (water-filled glove)	Ni=8 Nc=8	Video-box training vs no training	EOM Time taken Errors number	Yes	EOM (√) Time taken (-) Errors number (-)
Pearson et al 2002[74, 77, 83]	RCT	Intracorporeal knot tying *in vitro* (box trainer)	Ni=8 Nc=9 43 Medical students	Video-box training vs no training (unstructured group)	Performance on 10 knot tying trials	Not stated	Time taken (√)
Taffinder et al 1998[77, 83]	RCT	Laparoscopic tasks *in vitro* (MIST-VR)	Ni=5 Nc=5	Video-box training vs no training	Performance	Not stated	Performance (√)
Video-box training vs no training –in vivo							
Traxer et al 2001[77, 83]	RCT	Laparoscopic skills *in vitro* (bench trainer) and *in vivo* (pig)	Ni=6 Nc=6 Urology residents	Video-box training vs no training	5 tasks on bench trainer Performance on porcine laparoscopic nephrectomy	Yes	Performance (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.6. Summary of the individual studies reviewed pertaining to video-box surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box training vs no training –in vivo							
Scott et al 2000[74, 76, 77, 83]	RCT	Laparoscopic cholecystectomy *in vitro* (video-box trainer) and *in vivo* (OR)	Ni=9 Nc=13	Video-box training vs no training	Performance	Yes	Performance *in vitro* (√) Performance *in vivo*: Overall performance (√) Respect of tissue (√) Instrument handling (√) Use of assistants (√) Knowledge of instruments or procedure (-) Time and motion (-) Flow of operation (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.6. Summary of the individual studies reviewed pertaining to video-box surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box training vs no training –in vivo							
Scott et al 1999[76]	RCT	Laparoscopic cholecystectomy *in vivo* (OR)	Ni=9 Nc=13	Video-box training vs no training	Overall performance Flow of operation		Performance(√) Flow of operation (-) Time and motion (√)
Fried et al 1999[74, 77, 83]	RCT	Laparoscopic performance *in vivo* (pig)	Ni=6 Nc=6 Junior surgical residents	Video-box training vs no training	Performance on 7 tasks	Not stated	Performance: 5 out of 7 tasks (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.7. Summary of the individual studies reviewed pertaining to video-box surgical training compared to standard training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box training vs standard training –in vitro							
Matsumoto et al 2002[77, 83]	RCT	Endourological skills *in vitro* (video-box trainer)	Ni=17 Nc=7 40 medical students	Video-box training vs didactic training (1hour instruction)	Performance scores Time taken	Yes	Performance (-) Time taken (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.8. Summary of the individual studies reviewed pertaining to video-box surgical training compared to simplified simulation

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Video-box *training vs simplified simulation –in vitro*							
Keyser et al 2002[77, 83]	Crossover RCT	Laparoscopic skills *in vitro* (video-box+ mirrored box)	N = 22	Video-box training vs simplified mirrored box training	Performance on 7 tasks	Not stated	Performance (-): 6 out of 7 tasks

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.9. Summary of the individual studies reviewed pertaining to VR surgical training compared to video-box training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs video-box training –in vitro							
Jordan et al 2000a[75, 77, 83]	RCT	Laparoscopic skills *in vitro* (box trainer)	N1=8 N2=8 N3=8 24 medical & science students	MIST-VR training vs video-box training (randomly alternating image) vs video-box training (normal image)	Performance (correct incisions)	Not stated	Performance (√)
Jordan et al 2001[74, 77, 83]	RCT	Laparoscopic skills *in vitro* (box trainer)	N1=8 N2=8 N3=8 32 medical & science students	MIST-VR training vs video-box training (Z-maze) vs video-box training (U-maze)	Performance (correct incisions)	Not stated	Performance (√)
Kothari et al 2002[75, 77, 83]	RCT	Laparoscopic skills *in vitro* (box trainer)	N1=13 N2=11 Medical students	MIST-VR training vs video-box training	Knot tying time	Not stated	Time (-)
Lehman et al 2005[77, 83]	RCT	Laparoscopic surgery *in vitro* (switched simulators)	N1= 16 N2=16 24 Medical students & 8 Surgeons	VEST VR training vs video-box training	Task completion time	No	Task completion time (-)
Munz et al 2004[75, 77]	RCT	Clip application *in vitro* (water-filled glove)	N1=8 N2=8	LapSim VR training vs video-box training	EOM Time taken Errors number	Yes	EOM (-) Time taken (-) Errors number (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.9. Summary of the individual studies reviewed pertaining to VR surgical training compared to video-box training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs video-box training –in vitro							
Pearson et al 2002[74, 75, 77, 83]	RCT	Intracorporeal knot tying *in vitro* (box trainer)	N1=10 N2=8 43 Medical students	MIST-VR training vs video-box training	Performance on 10 knot tying trials	Not stated	Time taken (-)
Torkington et al 2001[74, 75, 77, 83]	RCT	Laparoscopic tasks *in vitro* (box trainer)	Ni=10 Nc=10 30 medical students	MIST-VR training vs box trainer surgical drills	Time taken to perform laparoscopic tasks	Not stated	Time taken (-)
Xia et al 2000[75]	RCT	Running sutures *in vitro* (video-box trainer)	41 participants with no laparoscopic experience	Karlsruhe VR training vs video-box training	No. of finished stitches in 0.5 hour, suture accuracy		No. of finished stitches (-) Accuracy (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.9. Summary of the individual studies reviewed pertaining to VR surgical training compared to video-box training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs video-box training –in vivo							
Hamilton et al 2002[75, 77, 83]	RCT	Laparoscopic cholecystectomy *in vitro* (MIST-VR and video-box) Laparoscopic cholecystectomy *in vivo* (OR)	N1=24 N2=25 Junior surgery residents N=19	MIST-VR training vs video-box training	Performance	Yes	Performance *in vitro*(√) Performance *in vivo* (√)
Madan et al 2007[75]	RCT	Laparoscopic skills *in vivo* (pig)	65 participants with no laparoscopic experience	MIST-VR training vs video-box training	Time Accuracy Subjective scores (0 to 100)	Yes	Time (-) Error score (-) Subjective scores (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.10. Summary of the individual studies reviewed comparing different types of video-box surgical training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Comparisons of different types of video-box training – in vitro							
Harold et al 2002[77, 83]	RCT	Laparoscopic suturing *in vitro* (bench model +video-box)	N1= 9 N2= 8 Surgical residents	Video-box training + additional instruction vs video-box training	Performance	Not stated	Suturing time (-) Needle placement accuracy (-) Suture strength (-)
Risucci et al 2001[77, 83]	RCT	Laparoscopic tasks *in vitro* (video-box trainer)	N1=7 N2=7	Video-box training + additional instruction vs video-box training	Performance on 2 tasks	Yes	Object passing errors (√) Other tasks (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.11. Summary of the individual studies reviewed pertaining to VR surgical training with a pre-test post-test design

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training pre-test post-test –in vitro							
Aggarwal et al 2006[8]	Prospective trial	Non-ostial left renal balloon angioplasty *in vitro*	N1 = 12 Inexperienced N2 = 8 Experienced Consultant vascular surgeons	VIST VR training	Procedure time Contrast volume Fluoroscopy time		Novice: Procedure time(√) Contrast volume(√) Fluoroscopy time(–) Experienced: No significant improvements
Dawson et al 2007[8]	Prospective trial	Peripheral angioplasty *in vitro*	N=9 Vascular fellows	SimSuite training + didactic tutorials	Procedure time Contrast volume Fluoroscopy time Time to treat complications No of balloons, wires, stents		Procedure time(√) Contrast volume(√) Fluoroscopy time(√) Time to treat complications (√) No of balloons, wires, stents(–)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.11. Summary of the individual studies reviewed pertaining to VR surgical training with a pre-test post-test design (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training pre-test post-test –in vitro							
Dayal et al 2004[8]	Prospective trial	Carotid artery stending (CAS) *in vitro*	N1 = 16 Novice N2 = 5 Experienced in CAS	VIST VR training	Procedure checklist score Procedure time Contrast volume Fluoroscopy time Subjective: Guide wire manipulation		Novice: Procedure score(√) Procedure time(√) Contrast volume(√) Fluoroscopy time(√) Guide wire manipulation (√) Experienced: No significant improvements
Mahmood & Darzi 2004[21]	Prospective trial	Colonoscopy skills *in vitro* (Immersion Medical simulator)	N = 26 Physicians	Immersion Medical VR training without feedback	Colonscopy performance over five attempts		Performance (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.11. Summary of the individual studies reviewed pertaining to VR surgical training with a pre-test post-test design (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training pre-test post-test –in vitro							
Neequaye et al 2006[8]	Prospective trial	Iliac & renal angioplasty *in vitro*	N=20 Surgical trainees	VIST VR training	Procedure time Contrast volume Fluoroscopy time Placement accuracy		Iliac angioplasty: Procedure time(√) Contrast volume(-) Fluoroscopy time(-) Placement accuracy(√) Renal angioplasty: Iliac angioplasty: Procedure time(√) Contrast volume(-) Fluoroscopy time(-) Placement accuracy(–)
O'Toole 1999[73, 74]	Prospective trial	Vascular anastomosis *in vitro* (simulator)	N = 12 Medical students	VR training	Suturing performance (7 parameters)		Performance(√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.11. Summary of the individual studies reviewed pertaining to VR surgical training with a pre-test post-test design (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training pre-test post-test –in vitro							
Patel et al 2006[8]	Prospective trial	Carotid artery stending (CAS) *in vitro*	N = 20 Interventional cardiologists	VIST VR training	Procedure time Fluoroscopy time Contrast volume Catheter handling error		Procedure time (√) Fluoroscopy time (√) Contrast volume (√) Catheter handling error (√)
VR training pre-test post-test –in vivo							
Clark et al 2005[21]	Prospective trial	Colonoscopy *in vitro* (simulator) and *in vivo* (bedside)	N = 13 PGy-1 & senior surgical residents	Simbionix VR training	Performance over 2 years		Examination efficiency (√) Absence of correlation to bedside-training cases
Prystowsky et al 1999[73]	Prospective trial	IV insertion *in vivo* (on another study participant)	N=51 37 1st year medical students 14 3rd year medical students	VR training	IV insertion success rate Performance Time to successful venous canulation		IV insertion success rate (-) Performance (-) Time: 1st years (-) 3rd years (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vitro							
Aggarwal et al 2007[75]	RCT	Laparoscopic cholocystectomy *in vitro* (porcine organs in video-box)	N= 20 participants with no laparoscopic experience	LapSim training vs no training	Objective structured assessment of technical skill (OSATS)	Yes	Time (√) No of movements (√)
Eversbusch & Grantcharov 2004[77]	RCT	Colonoscopy *in vitro* (computer simulator)	Ni=10 Nc=10	GI Mentor II VR training vs no training	Performance Time taken	Not stated	Performance (√) Time taken (√)
Ferlitsch et al 2002[21, 83]	RCT	Endoscopy skills *in vitro* (VR simulator)	Ni=7 Nc=6 Medical students N = 11 experts	GI Mentor VR training vs no training	Virtual endoscopy performance	Not stated	Performance (√)
Gallagher et al 1999[73, 75, 77]	RCT	Laparoscopic incisions *in vitro* (box trainer)	Ni=8 Nc=8	MIST-VR training vs no training	Correct incisions number	Not stated	Correct incisions no (√)
Hsu et al 2004[8]	RCT	Carotid artery stending (CAS) *in vitro*	N = 29 16 Untrained & 13 Advanced	VIST VR training vs no training	Time to successful completion		Improvement in time to successful completion (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)
*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vitro							
Jordan et al 2001[74, 77, 83]	RCT	Laparoscopic cutting skills *in vitro* (box trainer)	Ni=8 Nc=8 Medical & science students	MIST-VR training vs no training	Laparoscopic cutting skills	Not stated	Laparoscopic cutting skills (√)
Munz et al 2004[75, 77]	RCT	Laparoscopic clip application *in vitro* (water-filled glove)	Ni=8 Nc=8	LapSim VR training vs no training	EOM Time taken Errors number	Yes	EOM (√) Time taken (-) Errors number (-)
Pearson et al 2002[74, 75, 77, 83]	RCT	Intracorporeal knot tying *in vitro* (box trainer)	Ni=10 Nc=9 43 Medical students	MIST-VR training vs no training (unstructured group)	Performance on 10 knot tying trials	Not stated	Time taken(√)
Tanoue et al 2005[75]	RCT	Laparoscopic suturing *in vitro* (Video-box trainer)	N=35 participants with no laparoscopic experience	MIST-VR training vs 30 min video instruction	Time Errors		Time (√) Error incidence (√)† †Statistical significance unknown
Torkington et al 2001[74]		Laparoscopic skills *in vitro* (MIST-VR)	N=13 Surgical trainees	MIST-VR training vs no training	Laparoscopic skills		Laparoscopic skills (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vitro							
Torkington et al 2001[74, 75, 77, 83]	RCT	Laparoscopic tasks *in vitro* (box trainer)	Ni=10 Nc=10 30 medical students	MIST-VR training vs no training	Speed Laparoscopic tasks	Not stated	Speed(√) Laparoscopic tasks (-)
Watterson et al 2002[77, 83]	RCT	Ureteroscopy *in vitro* (URO mentor)	Ni=10 Nc=10 Medical students	URO Mentor VR training vs no training	Distal calculus performance out of 25	Yes	Performance (√)
Wilhelm et al 2002[77, 83]	RCT	Endoscopy skill *in vitro* (URO mentor)	Ni=11 Nc=10	URO Mentor VR training vs no training	Proximal ureteral calculus performance out of 25	Not stated	Performance (√)
VR training vs no training –in vivo							
Ahlberg et al 2002[71, 74, 75, 77, 83]	RCT	Laparoscopic appendectomy *in vivo* (pig)	Ni=14 Nc=15 Medical students	MIST-VR training vs no training	Performance score of 30	Yes	Performance (-)
Ahlberg et al 2005[76]	RCT	Colonoscopy *in vivo* (patients)	Ni = 6 Nc = 6	AccuTouch VR training vs no simulation training	Time to reach the cecum Success rate (%) Patient-assessed discomfort		Time taken (√) Success rate (√) Patient-assessed discomfort (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vivo							
Ahlberg et al 2007[75]	RCT	Laparoscopic cholecystectomy *in vivo* (10 patients)	N= 13 participants with limited laparoscopic experience	Lapsim VR training vs no training	Subjective error score	Yes	Error score (√)
Andreatta et al 2006[83]	RCT	Laparoscopic performance *in vivo* (pig)	N=19 Surgical interns	LapMentor VR training vs no training	Performance	Yes	Navigation speed (√) Accuracy (√)
Bensalah et al 2007[75]	RCT	Laparoscopic cholecystectomy *in vitro* (VR model) Laparoscopic nephrectomy *in vivo* (porcine model)	N= 32 participants with no laparoscopic experience	LapMentor training vs no training	OSATS	Yes	Composite score *in vitro* (√) Composite score *in vivo* (√)
Chaer et al 2006[8]	RCT	Endovascular skills *in vivo* (patients)	N=20 General surgery residents	VIST VR training vs no simulator training	Checklist (18 steps) Global rating scale		Checklist (√) Global rating (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vivo							
Cohen et al 2006[21, 76]	Multi-centre RCT	Colonoscopy *in vivo* (patients)	Ni=23 Nc=22 Gastroenterology fellows	GI Mentor VR training vs no training	Competence (ability of reaching the splenic flexure and cecum independently, note abnormalities, recognise pathology correctly) Patient discomfort		Cases 1-20: Competence (–) Cases 21-80: Competence (√) Overall: Performance (√) Patient discomfort (-)
Cosman et al 2007[75]	RCT	Clipping and division of blood vessel *in vivo* (humans)	N=10 participants with limited laparoscopic experience	Lapsim VR training vs no training	Time Subjective error score	Yes	Time (-) Error score (√)
Di Giulio et al 2004[83]	RCT	Upper endoscopy *in vivo* (20 patients)	N = 22 Gastroenterology fellows	GI Mentor VR training vs no training	Skills performance	No	No of complete procedures (√) Less assistance required(√)
Ganai et al 2007[75]	RCT	Angled telescope skills *in vivo* (pig)	N= 20 participants with no laparoscopic experience	Endotower training vs no training	Time Subjective error score	Yes	Time (-) Error score (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vivo							
Grantcharov et al 2004 [71, 75-77, 83]	RCT	Laparoscopic cholecystectomy *in vivo* (1 patient)	Ni=8 Nc=8 Surgical trainees	MIST-VR training vs no training	Accuracy Operation times	Yes	Accuracy (√) Operation times (-)
Hyltander et al 2002[71, 75, 77, 83]	RCT	Laparoscopic navigation *in vivo* (pig)	Ni=12 Nc=12 Medical students	LapSim VR training vs no training	Performance Time taken	Yes	Performance(√) Time (√)
Madan et al 2007[75]	RCT	Laparoscopic skills *in vivo* (pig)	N= 65 participants with no laparoscopic experience	MIST-VR training vs no training	Time Accuracy Subjective scores (0 to 100)	Yes	Time (-) Error score (-) Subjective scores (√)
McClusky et al 2004[75]	RCT	Laparoscopic cholecystectomy *in vivo* (humans)	N= 12 participants with limited laparoscopic experience	MIST-VR training vs no additional training	Time Subjective error score	Yes	Time (√) Error score (√)
Schijven et al 2005[76]	Non-randomised comparative study	Laparoscopic cholecystectomy *in vivo* (patients)	Ni=10 Nc=10	Laparoscopic cholecystectomy course including VR sessions vs no training course	Flow of movement (scale 0-5)	Yes	Flow of movement (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.12. Summary of the individual studies reviewed pertaining to VR surgical training compared to no training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs no training –in vivo							
Tuggy 1998[21, 76]	RCT	Flexible sigmoidoscopy *in vivo* (healthy volunteers)	Ni=5 Nc=5 Family medicine residents	Gastro-Sim VR training vs no training	Hand eye skills performance	Yes	Insertion time (√) Directional errors (√) Examination time (√)
Youngblood et al 2005[77, 83]	RCT	Laparoscopic performance *in vivo* (pig)	Ni=17 Nc=13 46 medical students	LapSim VR training vs no training	Laparoscopic performance scores	Yes	Performance (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)
*Blank cells represent information not found in the reviews

Table 3.13. Summary of the individual studies reviewed pertaining to VR surgical training compared to standard training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs standard training – in vitro							
Bowyer et al 2005[83]	RCT	IV cannulation *in vitro* (simulated arm)	N= 34 Medical students	CathSim VR training vs Virtual IV VR training vs plastic arm vs *in vivo* practice	Performance	Not stated	Virtual IV group: Performance (√)
Pearson et al 2002[74, 77, 83]	RCT	Intracorporeal knot tying *in vitro* (box trainer)	N1=10 N2=8 N3=8 43 Medical students	MIST-VR training vs self-practice vs didactic instruction	Performance on 10 knot tying trials	Not stated	Time taken (-)
VR training vs standard training – in vivo							
Chang et al 2002[82]	RCT	Venous cannulation *in vivo* (patient)	N=28 Nurses	CathSim VR training vs traditional training (plastic arm)	Cannulattion success Checklist-rated performance		Success (X)† †Statistical significance unclear Performance (-)
Engum et al 2003[82, 83]	RCT	IV catheter placement *in vivo* (simulated patient)	N = 163 Medical & nursing students	CathSim VR training vs traditional training (with manikin)	Checklist-rated performance Knowledge in paper test	Not stated	Performance (-) Knowledge (X)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.13. Summary of the individual studies reviewed pertaining to VR surgical training compared to standard training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs standard training – in vivo							
Gerson & Van Dam 2003[21, 71, 76, 83]	RCT	Flexible sigmoidoscopy *in vivo* (5 patients)	Ni= 9 Nc=7 Internal medicine residents	Immersion Medical VR simulator training vs bedside training	Skills performance Patient satisfaction Patient discomfort	No	Performance (X) Time (-) Patient satisfaction(-) Patient discomfort (-)
Ost et al 2001[71, 83]	RCT	Bronchoscopy *in vivo* (2 patients)	Ni= 3 Nc=3 Pulmonary fellows	AccuTouch VR training vs bedside training	Performance	Yes	Performance (√)
Peugnet et al 1998[73]		Retinal photocoagulation *in vivo* (patient)	Residents	VR training (14.8 sessions) vs bedside training (11.25 sessions)	Efficiency (intensity, size and distance of laser spots, number of spots, session duration)	Yes	Efficiency (-)
Rowe & Cohen 2002[71, 84]	RCT	Fiber optic Intubation *in vivo* (patients)	N1= 12 N2= 8 Paediatric residents	AccuTouch VR training vs traditional training	Task completion time Errors		Error score (√) Task completion time (√)
Sedlack et al 2004[21, 76]	RCT	Flexible sigmoidoscopy *in vivo* (patients)	Ni=19 Nc=19 Internal medicine residents	AccuTouch VR simulator + bedside training vs bedside training	Procedural skills performance Patient discomfort score (scale 1-10)		Performance (-) Patient discomfort (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

*Blank cells represent information not found in the reviews

Table 3.13. Summary of the individual studies reviewed pertaining to VR surgical training compared to standard training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
VR training vs standard training – in vivo							
Sedlack & Kolars 2004[21, 76]	RCT	Colonoscopy *in vivo* (15 patients)	Ni=4 Nc=4 Gastroenterology fellows	AccuTouch VR training +bedside training vs bedside training	Colonoscopy performance Insertion time Patient discomfort score (scale 1-10)	No	Initial performance (√) Insertion time (-) Patient discomfort for colonoscopies 1-15 (√)
Seymour et al 2002[71, 74-77, 83]	RCT	Laparoscopic cholecystectomy *in vivo* (OR)	Ni=8 Nc=8 Junior surgical residents	MIST-VR + standard training vs standard training (video? not clearly defined)	Laparoscopic cholecystectomy performance	Yes	Errors number (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)
*Blank cells represent information not found in the reviews

Table 3.14. Summary of the individual studies reviewed comparing different types of VR surgical training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Comparisons of different types of VR training – in vitro							
Ali et al 2002[77]	RCT	Surgical performance *in vitro* (MIST-VR)	N1=13 N2=14 Novice	MIST-VR training: Medium vs easy level training	Performance scores	Not stated	Medium level training: Performance (√)
Hassan et al 2005[75]	RCT	Clip application task *in vitro*	N= 14 participants with no laparoscopic experience	Lapsim training: Easy vs difficult level training	Speed Blood loss	Not stated	Easy level training: Speed (√) Blood loss (√)
Lagrana et al 1997[73]		Sub-surface liver tumour palpation *in vitro* (VR simulator)	32 non-medical students	5min VR training vs 1.5 min VR training	Location and differentiation of hard and soft tumours Time taken		Time (-) Differentiation of tumours (-) Tumour location (-)
Mackay et al 2002[75, 77, 83]	RCT	Laparoscopic skills *in vitro* (MIST-VR)	N1=14 N2=14 N3=13 Students	MIST-VR training: 20 min massed practice vs 20 min distributed (in 5 min blocks) practice vs 15 min distributed practice	Performance scores	Not stated	20 min distributed practice: Performance (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.14. Summary of the individual studies reviewed comparing different types of VR surgical training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Comparisons of different types of VR training – in vivo							
Verdaasdonk 2008[75]	RCT	Surgical knot *in vivo* (pig)	N= 20 participants with no laparoscopic experience	Basic+ knot tying module Simendo® training vs basic module Simendo® training	Time for driving a needle through tissue Time to tie a knot Error scores	Yes	Time for driving a needle through tissue (-) Time to tie a knot (-) Objective error score (√) Subjective error score (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.15. Summary of the individual studies reviewed pertaining to ex-vivo model surgical training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Ex-vivo model training pre-test post-test							
Maiss et al 2005[21]	Prospective study	Endoscopic haemostasis skills *in vitro* (Compact EASIE simulator)	N=32 Gastroenterology fellows	CompactEASIE model training	Subjective performance (precision in argon plasma coagulator use, variceal ligation, injection & coagulation, hemoclip application)	No	Performance (√)
Ex-vivo model training vs standard training							
Hochberger et al 2005[21, 83]	RCT	Endoscopy skills *in vitro* (simulator) and *in vivo* (clinical procedures)	N= 23 Gastroenterology fellows	CompactEASIE model training + clinical training vs clinical training only	Performance in 4 skills (precision in using coagulator, variceal ligation, injection and coagulation, hemoclip application) Clinical hemostatic procedures outcomes	Yes	Performance in 4 skills (√) Clinical procedures success rate (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

Table 3.16. Summary of the individual studies reviewed pertaining to cadaver surgical training compared to standard training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Cadaver training vs standard training –in vitro							
Anastakis et al 1999[77, 83]	RCT	Surgical tasks *in vitro* (cadaver)	N = 23 Junior surgical residents	Cadaver training vs standard training (independent learning from manual)	Performance of 6 surgical procedures	Yes	Performance (√)
Cadaver training vs no training - in vivo							
Martin et al 2003[82]	Retrospective with historical controls	Central venous line insertion *in vivo* (patients)	N= 105 Medical students	Cadaver training vs no training	Rate of pneumothorax		Rate of pneumothorax (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

APPENDIX 3.3. REVIEW OF REVIEWS DATA: INDIVIDUAL STUDIES PERTAINING TO MEDICINE (TABLES 3.18 – 3.25)

Table 3.18. Summary of the individual studies reviewed pertaining to model medical training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Inanimate model training vs no training -in vitro							
Aliabadi-Wahle et al 2000[78]	RCT	Breast examination *in vitro* (breast models)	N= 30 Medical students	Model training vs no training	Ability to detect lumps		Lump detection (√)
Davies &Gould 2000[79]	Quasi-experimental	CPR skills *in vitro* (mannequin)	Student nurses Sample size not reported	Self-instruction skillmeter mannequin retraining vs no retraining	CPR performance		Performance (√)
Madan et al 2002[78]	Clinical trial	Breast examination *in vitro* (breast models)	N= 47 Medical students	Video + model training vs no training	Ability to detect lumps		Lump detection (√)
Inanimate model training vs standard training – in vivo							
Hosking et al 1998[83]	RCT	Laryngoscopy + intubation *in vitro* (MCQ) and *in vivo* (OR)	N = 46 Medical students	METI mannequin training vs standard training (demonstration)	Self-rated performance on intubations over 3 week clinical placement Knowledge test	No	Intubation success rate (-) Knowledge (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.18. Summary of the individual studies reviewed pertaining to model medical training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Inanimate model training vs standard training – in vivo							
Stewart et al 1984[82]	Prospective trial	Intubation *in vivo* (patients)	N = 146 Paramedics	Mannequin training vs mannequin + animal-based training vs mannequin + animal-based + OR training vs standard training	Performance in the field over 27 months		Success (-) Time taken (-) Complications (-)
Inanimate model training vs other forms of simulation training -in vitro							
Gerling et at 2003[78]	Clinical trial	Breast examination *in vitro* (breast models)	N=48 Medical students	Dynamic model training vs standard model training	Ability to detect lumps		Lump detection (√)
Kovacs et al 2000[82]	RCT	Intubation *in vitro*	N=84 Health science students	Mannequin training (control) vs mannequin training + periodic feedback vs mannequin training + independent mannequin practice + intermittent feedback	Performance checklist (52 points) at 0, 16,25 and 40 weeks		mannequin training + independent practice + intermittent feedback vs control: Performance (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.18. Summary of the individual studies reviewed pertaining to model medical training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Inanimate model training vs other forms of simulation training -in vivo							
Levitan et al 2001[82]	Prospective trial with retrospective controls	Intubation in vivo (patients)	Ni=36 Nc=113 Paramedics	Airway training programme (Textbook + lectures+ mannequin practice) vs Airway training programme + 26 min laryngoscopy video	Performance (success rates)		Airway training programme plus video: First attempt success rates (√) Overall success rates (√)
Stewart et al 1984[82]	Prospective trial	Intubation in vivo (patients)	N = 146 Paramedics	Mannequin training vs mannequin + animal-based training vs mannequin + animal-based + OR training vs standard training	Performance in the field over 27 months		Success (-) Time taken (-) Complications (-)
Stratton 1991[82]	Prospective trial	Intubations in vivo (patients)	Ni=30 Nc=30 Paramedics	Mannequin training vs mannequin+ cadaver training	Successful intubation in the field		Success rate (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.19. Summary of the individual studies reviewed pertaining to computer-based simulation medical training with a pre-test post-test design

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Computer Simulation training pre-test post-test							
Wik et al 2002[79, 84]	Cohort study	CPR skills *in vitro* (Voice-activation mannequin - VAM)	N= 35 Nonmedical employees	VAM training (20 min) without an instructor	CPR skills performance		Immediate performance (√) Performance at 6 months (-)
Wik et al 2001[79]		CPR skills *in vitro* (VAM)		VAM training without an instructor	CPR skills performance		Immediate performance in the presence of audio feedback (√)
Harrell et al 1990[81]	Pre-test, post test	Heart sounds identification	N= 37 Graduate nursing students	Heart sound simulator teaching	Knowledge		Knowledge (√)
Harrell et al 1990[81]	Pre-test, post test	Heart sounds identification	N= 40 Registered nurses	Heart sound simulator teaching	Knowledge		Knowledge (√)
Champagne et al 1989[81]	Pre-test, post-test with unspecified control group	Cardiac auscultation *in vitro* (simulator)	N= 37 Registered nurses	Heart sound simulator training	Identification of abnormal cardiologic physical findings		Heart sounds identification (√)
Garfield et al 1989[81]	Pre-test, post-test with crossover	Anaesthetic uptake and distribution *in vitro* (written tests)	N=16 Anaesthesia residents	Individualised GasMan simulator teaching	22-object written tests		Knowledge (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.19. Summary of the individual studies reviewed pertaining to computer-based simulation medical training with a pre-test post-test design (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Computer Simulation training pre-test post-test							
Howard 1987[81]	Pre-test, post-test	Cardiovascular arrest *in vitro* (written test)	N= 97 Registered nurses	Computer simulation for continuing education	Knowledge of cardiovascular arrest		Knowledge (√)
Woolliscroft et al 1987[81]	Pre-test, post-test	Ausculatory, tactile, visual findings *in vitro* (Harvey)	N=203 Medical students	Harvey patient simulator training	Cardiology skills		Skills (√)
Gordon et al 1980[81]	Pre-test, post-test	Practical bedside cardiologic skills Knowledge *in vitro* (multiple-choice questions (MCQ) exam)	N = 34 Senior medical students	Cardiologic patient simulator training	Bedside examination skills Cognitive information		Practical skills (√) Knowledge (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.20. Summary of the individual studies reviewed pertaining to computer-based simulation medical training compared to no training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Computer Simulation training vs no training –in vitro							
Gilbart et al 2000[83]	RCT	Trauma management *in vitro* (OSCE)	N=179 Medical students	Computer-based simulation vs no training	Performance on OSCE	Not stated	Performance (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.21. Summary of the individual studies reviewed pertaining to computer-based simulation medical training compared to standard training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Computer Simulation training vs other forms of training-in vitro							
Wayne et al 2005[83]	RCT	Advance cardiac life support skills (ACLS) *in vitro* (simulator)	N=38 Internal medicine residents	HPS simulator training vs standard clinical experience	ACLS performance	Yes	Performance (√)
Curran et al 2004[83]	RCT	Neonatal resuscitation *in vitro* (simulator)	N= 31 Medical students	ANAKIN computer-mediated mannequin vs video training	Skills performance Knowledge	Not stated	At 8 months: Performance (-) Knowledge (–)
Morgan et al 2002[83]	RCT	Anaesthesia management in critical event *in vitro* (simulator)	N = 144 Medical students	Simulator training vs video training	Management performance on anaesthesia critical event scenario	Not stated	Performance (-)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.21. Summary of the individual studies reviewed pertaining to computer-based simulation medical training compared to standard training (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Computer Simulation training vs other forms of training-in vitro							
Treadwell et al 2002[82]	Prospective trial with retrospective controls	Intubation *in vitro*	N=94 Medical students	Computer-based (CD ROM + lab-based) training vs traditional training	Performance checklist		Performance (X)† †Statistical significance unclear
Gilbart et al 2000[81, 83]	RCT	Trauma management *in vitro* (OSCE)	N=107 179 medical students	Computer-based simulation vs seminar-based teaching	Performance on OSCE	Not stated	Performance (-)
Schwid et al 1999[79]	RCT	Cardiac arrest management *in vitro* (simulation test)	N=45 Anaesthetists	Computer-based ACLS simulation programme vs textbook study	Cardiac arrest simulation performance		Performance (√)
Ewy et al 1987[81, 84]	Cohort	Cardiology knowledge and skills *in vitro* (MCQ exam, Cardiologic patient simulator) and *in vivo* (patients)	N= 208 4th year medical students	Cardiology elective including cardiologic patient simulator vs cardiology elective without the simulator	Cardiology knowledge Cardiology skills		Knowledge (√) Cardiology skills *in vitro* & *in vivo* (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.22. Summary of the individual studies reviewed comparing different types of computer-based simulation medical training

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Comparisons of different types of computer simulation training – in vitro							
Wik et al 2002[79, 84]	Cohort study	CPR skills *in vitro* (VAM)	N= 35 Nonmedical employees	50 min VAM distributed training vs 20 min VAM training	Retention of CPR skills at 6 months		CPR skills retention (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)
* Blank cells represent information not found in the reviews

Table 3.23. Summary of the individual studies reviewed pertaining to simulated patients medical training with a pre-test post-test design

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients pre-test post test –in vitro							
Cushing & Jones 1995[78]	Descriptive study	Breaking bad news *in vitro* (questionnaire)	N=231 Medical students	Simulated patients +role play +video course	Knowledge		Knowledge (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)
* Blank cells represent information not found in the reviews

Table 3.24. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to no intervention

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients vs no intervention							
Haist et al 2004[80]	Appears non-randomised, no baseline assessment	Sexual history enquiry and HIV counselling	Medical students Sample size not reported	Simulated patient workshop training vs no training	Sexual health and HIV counselling behaviours	Unclear	Sexual health and HIV counselling behaviours (√)
Haist et al 2003[80]	Non-randomised, no baseline assessment but groups analysed for demographic characteristic equivalence	Domestic violence consultation	Medical students Sample size not reported	Simulated patient workshop training vs no such training	Clinical performance examination checklist	Not stated	Domestic violence item scores (√) Interpersonal skills (-)
Fallowfield et al 2002[80]	RCT	Communication skills for oncologists *in vivo* (patients)	Oncologists Sample size not reported	Course (including simulated patient consultations) with written feedback after vs course alone vs written feedback alone vs no intervention	Validated outcome measure of communication skills	Yes	Course attendance: Communication skills (√)
Kruijver et al 2001[80]		Patient centred communication *in vitro* (simulated patients)	Cancer nurses Sample size not reported	Simulated patient training vs no training	Patient centred communication skills		Open-ended, psychosocial questions (√) Affective talk (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.24. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to no intervention (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients vs no intervention							
Colletti et al 2001[78, 80]	Clinical trial, no baseline assessment	Breaking bad news *in vitro* (simulated patient)	N=38 Medical students	Simulated patient training vs no training	27-item measure of communications skills	Not stated	Communication skills score (√)
Stillman et al 1997[80]		Clinical interviewing skills	3rd, 4th and 5th year medical students Sample size not reported	Simulated patient training (1993 curriculum) vs no such training (1992 curriculum)	Adult and paediatric interviewing skills		Adult and paediatric interviewing skills (√)
Johnson & Kopp 1996[80]	No baseline assessment	Dental consultation *in vitro* (simulated patient)	1st year and 2nd year dental students Sample size not reported	Simulated patient training (1st years) vs previous real patient consultation experience (2nd years)	Record keeping, examination and communication skills checklist	Unclear	Record keeping (√) Examination (√) Communication skills (-)
Rabin et al 1994[80]	RCT, no baseline assessment	Office-based prevention practices for sexually transmitted diseases	General practitioners Sample size not reported	Simulated patient + mailed educational materials training vs no training	HIV risk assessment and counselling skills		Condom use and risky sex practices questions (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.24. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to no intervention (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients vs no intervention							
Koerber et al 2003[80]	RCT	Brief motivational interviewing for smoking-cessation counselling *in vivo*? (patients)	N=22 Dental students	Simulated patient course training vs no training	Brief motivational interviewing techniques displayed in practice	Unclear	Brief motivational interviewing techniques (√) Active involvement of patients (√)
Cornuz et al 2002[80]	RCT	Smoking-cessation counselling *in vivo* (patients)	Medical residents Sample size not reported	Simulated patient + role-play training vs didactic lecture	Smoking cessation counselling		Smoking cessation counselling skills (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information not found in the reviews

Table 3.25. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to other training methods

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients vs other methods of training							
Zraick et al 2003[80]	Randomised trial, no baseline assessment	Communication skills in speech/ language pathology *in vitro* (OSCE)	N=18 Students in speech- language pathology	Simulated patient training vs didactic lectures	Interpersonal and communication skills in OSCE	Not stated	Communication skills (–)
Blue et al 1998[80]	No baseline assessment	Breast examination	Students Sample size not reported	Simulated patient training vs didactic lectures	Communication skills	Not stated	Communication skills (√)
Madan et al 1998[80]		HIV risk assessment	N1=6 N2=6 Medical residents	Simulated patient training vs didactic lectures	HIV risk assessment skills		Communication skills (√)
Papadakis et al 1997[80]	Randomised trial	Smoking cessation *in vitro* (simulated patient)	1st year medical students Sample size not reported	Simulated patient training vs role-play training	Cognitive and communication skills rating form	Yes	Consultation skills (-)
Chalabian et al 1996[78]	Cohort study	Breast examination *in vitro* (OSCE)	N=120 House officers & medical students	Structured clinical instruction module vs normal teaching	Breast examination skills score		Breast examination skills (√)
Heard et al 1995[78]	Clinical trial	Breast examination *in vitro* (knowledge test + OSCE)	N=144 Medical students	Teaching by standardised patients vs normal teaching	Knowledge test Breast examination skills in OSCE		Knowledge (-) Breast examination skills (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information missing from the reviews

Table 3.25. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to other training methods (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Simulated patients vs other methods of training							
Campbell et al 1994[78]	RCT	Breast examination *in vitro* (models)	N = 54 Medical students	Teaching from standardised patients vs normal teaching	Ability to detect lumps		Lump detection: Sensitivity (√) Specificity (X)
Rabin et al 1994[80]	RCT, no baseline assessment	Office-based prevention practices for sexually transmitted diseases	General practitioners Sample size not reported	Simulated patient + mailed educational materials vs mailed educational materials	HIV risk assessment and counselling skills		Condom use and risky sex practices questions (√)
Pilgrim et al 1993[78]	RCT	Breast examination *in vitro* (models	N = 156 Medical students	Teaching from standardised patients+ video + lecture vs video+ lecture	Ability to detect lumps		Lump detection (√)
Fallowfield et al 2002[80]	RCT	Communication skills for oncologists *in vivo* (patients)	Oncologists Sample size not reported	Course (including simulated patient consultations) with written feedback after vs course alone vs written feedback alone vs no intervention	Validated outcome measure of communication skills	Yes	Course attendance: Communication skills (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information missing from the reviews

Table 3.25. Summary of the individual studies reviewed pertaining to simulated patients medical training compared to other training methods (continued)

Study	Design	Procedure	Subjects and sample size	Comparators	Outcome measures	Assessment blinding	Results
Comparisons of different types of simulated patients training – in vitro							
Kleinman et al 1996[80]	No baseline assessment	Pelvic examination *in vitro* (simulated patient)	Medical students Sample size not reported	Laywoman-simulated patient teaching 3rd years vs physician teaching 2nd years using a simulated patient	Pelvic examination skills	No	Technical skills (–) Interpersonal skills (√)

(√ = Intervention superior, X = Intervention worse, - = no statistically significant difference)

* Blank cells represent information missing from the reviews

APPENDIX 4.1. THE LMA STUDY BASELINE QUESTIONNAIRE

UNIVERSITY OF BIRMINGHAM

THE EFFECTS OF LARYNGEAL MASK AIRWAY (LMA) PASSAGE SIMULATION TRAINING ON THE ACQUISITION OF CLINICAL SKILLS

Baseline Assessment Participant Questionnaire

Version 6, 31/03/06

We would be very grateful if you could answer the following questions on your demographics and previous experience in performing LMA passage. All information collected will remain confidential and no links to participants will be made. All data will be held securely in line with the Data Protection Act.

PLEASE PRINT:

1. Name: ……...Student ID……..

2. Sex (Please tick): Male Female

3. Date of Birth ……/……/……

4. City/Country of Birth……..

5. Ethnicity:..……...

6. What is your parents' highest educational qualification? (Please tick the appropriate)

	Father	Mother
O' levels or equivalent		
A' levels or equivalent		
Bachelor's Degree		
Master's Degree		
Ph.D.		
Not known		
Other (please specify):	……..	……

7. What were your A-level subjects and grades? (Please fill in the table)

Subject	**Grade**	**Subject**	**Grade**	**Subject**	**Grade**	**Subject**	**Grade**
Biology		Chemistry					

8. Have you had your ARICM clinical skills mannequin training?
   Yes No

If yes, how many times did you practice LMA placement on a mannequin?

_____ Times

9. Had you ever been taught LMA placement prior to your ARICM module?

   Yes No

If yes, please briefly state the training method(s) used:.....

10. Have you ever performed LMA placement on a patient?

    Yes No If yes, on how many patients? _____ Patients

11. Where are you currently placed? ICU Operating theatre

THANK YOU

APPENDIX 4.2. THE LMA STUDY BASELINE ASSESSMENT PRO-FORMA

THE EFFECTS OF LARYNGEAL MASK AIRWAY PASSAGE SIMULATION TRAINING ON THE ACQUISITION OF CLINICAL SKILLS

Baseline Assessment Pro-Forma **version 9, 30/03/06**

Assessor's name ……

1. Number of insertion attempts _____

2. Time taken from the point at which the student touched the mannequin to the time they were able to ventilate the mannequin

_____ sec

3. Insertion success (Please tick the appropriate)

The student:

stood behind the head of the mannequin	Yes ☐	No ☐
inserted the cuff deflated (or partially inflated)	Yes ☐	No ☐
inserted the LMA with the correct side facing the airway	Yes ☐	No ☐
inflated the cuff upon insertion	Yes ☐	No ☐
LMA placement led to adequate chest movement	Yes ☐	No ☐

4. How would you rate the student's handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Rough	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

5. How would you rate the student's overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

APPENDIX 4.3. THE LMA STUDY CLINICAL PRACTICE ASSESSMENT FORM

UNIVERSITY OF BIRMINGHAM

STUDY TITLE: THE EFFECTS OF LARYNGEAL MASK AIRWAY PASSAGE SIMULATION TRAINING ON THE ACQUISITION OF CLINICAL SKILLS

CLINICAL SUPERVISOR ASSESSMENT PRO-FORMA version 11, 31/03/06

The student who forwarded this pro- forma to you is taking part in the abovementioned study. This study aims to investigate how simulation medical training translates into real life clinical practice. Your cooperation would be greatly appreciated. Please take a minute to read through this pro-forma and record your assessment of the student's performance. Once the assessment is completed, please seal the pro-forma in the provided envelope and return it to the participant. Please do not disclose the contents of this pro-forma to the participant.

Clinical supervisor's name:..……..……...
Participant's name:**……**
LMA placement attempt no … **Date……**

1. Number of insertion attempts: _____

2. Time taken from the point at which the student touched the patient to the point they were able to successfully ventilate the patient

_____ sec If not possible to time please tick an estimate ≤ 40sec ☐ >40sec ☐

3. If the attempt to perform LMA passage was abandoned before the student was able to attempt ventilation or they were unable to ventilate successfully please state briefly the reason:

……
Other comments:**……..**

4. Ventilation success (Please tick the appropriate):

LMA placement led to an adequate seal ☐ Yes ☐ No

LMA placement led to easy assisted ventilation ☐ Yes ☐ No

LMA placement led to adequate chest movement ☐ Yes ☐ No

LMA placement led to a normal capnographic curve ☐ Yes ☐ No ☐ N/A

5. How would you rate the student's handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Poor	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

6. How would you rate the student's overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

This pro-forma is now complete. Thank you.

APPENDIX 4.4. THE LMA STUDY CLINICAL PRACTICE SELF-ASSESSMENT QUESTIONNAIRE

UNIVERSITY OF BIRMINGHAM

THE EFFECTS OF LARYNGEAL MASK AIRWAY (LMA) PASSAGE SIMULATION TRAINING ON THE ACQUISITION OF CLINICAL SKILLS

Clinical Practice Participant Questionnaire

Version 6, 31/03/06

We would be very grateful if you could gradually answer the following questions on each of your first four consecutive attempts of performing LMA passage in patients as you progress through your ARICM placement. Once completed, please return the questionnaire along with your clinical supervisors' assessment pro-formas via post using the stamped envelope, which has been provided to you. All information collected will remain confidential and no links to participants will be made. All data will be held securely in line with the Data Protection Act.

Clinical Practice Participant Questionnaire v6 31/03/06

Patient no 1

1. How would you rate your handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Rough	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

2. How would you rate your overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

Patient no 2

3. How would you rate your handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Rough	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

4. How would you rate your overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

Patient no 3

5. How would you rate your handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Rough	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

6. How would you rate your overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

Patient no 4

7. How would you rate your handling of the LMA during the insertion? (Please circle the appropriate)

Extremely Rough	Below Average	Average	Above Average	Extremely Smooth
1	2	3	4	5

8. How would you rate your overall success of LMA passage? (Please circle the appropriate)

Extremely poor	Below Average	Average	Above Average	Excellent
1	2	3	4	5

9. The LMA placement mannequin training that I received (study intervention training, if applicable, and ARICM) was helpful (please circle the appropriate):

Strongly disagree	Disagree	Undecided	Agree	Strongly agree	Did not receive any mannequin training
1	2	3	4	5	N/A

Please explain further (Please PRINT):

10. If you received your ARICM mannequin training after the start of this study (13th February), please indicate the number of times that you practiced LMA placement on the mannequin as part of your ARICM module

_____times

11. Name: Student ID:

12. Thank you for taking the time to fill in this questionnaire. If you have any additional comments to make please use this space (Please PRINT):

APPENDIX 4.5. THE LMA STUDY PARTICIPANT INSTRUCTIONS WITH REGARDS TO CLINICAL PRACTICE

The LMA Study - Clinical practice instructions

In order to investigate how your LMA simulation training translates into clinical practice we seek to obtain a written record of the first four consecutive times that you perform LMA placement on patients during your ARICM practice. Thus, we have included four assessment pro-formas in four sealed envelopes for you to carry with you and forward to your clinical supervisor each time you are about to perform an LMA insertion on a patient.

- Please forward one pro-forma each time before performing an LMA placement and ask the person supervising you to fill it in. Upon completion, they should return it back to you in a sealed envelope.
- Once all four pro-formas have been completed, please seal them along with your Clinical Practice Questionnaire in the provided envelope and drop it in your nearest mailbox. Alternatively, please contact Niki Laiou to collect these from you.
- Please do not reveal to your clinical supervisor whether you have received additional simulation training or not as this could affect their assessment.
- Your clinical supervisor will also have been asked not to disclose to you the contents of the assessment pro-formas in an effort to minimize the effects of the assessment on your performance.
- Should for some reason your clinical supervisors are unable to complete some part or all of an assessment pro-forma, they should return it to you to forward back to us and you should proceed with the next one as if that assessment had been completed. We are only interested on how you perform on your first four consecutive LMA placements in patients rather than any four LMA placements.
- The numbers on the four sealed envelopes denote the order in which the pro-formas should be filled in by your clinical supervisors. Should there be any changes in the order they are filled in, please mark these down on the envelopes returned to you.

Your help in this final stage of the study is of vital importance and would be greatly appreciated.

APPENDIX 5.1. THE MOLTI SCENARIOS

Scenario 1 (Baseline Assessment)

F1 Doctor in A&E
This is a 28 year old woman with known mild asthma for 10 years, no previous hospital admissions. Increasing shortness of breath for last 12 hours. Now feels dreadful and unable to talk. ECG, pulse oximeter and blood pressure monitoring is attached.

(Tension pneumothorax on chest X-ray (CXR) if requested)

Scenario 2

F1 A&E
This is a 23 year old man with no previous hospital admissions. Fell downstairs last week at a party, banged his head and bruised his left lower ribs, no loss of consciousness, no other injuries. Now feels unwell and short of breath. He has and ECG, a pulse oximeter and non-invasive blood pressure monitoring attached.

(Ruptured spleen, hidden blood loss ++, resuscitation and cross matching blood if requested)

Scenario 3

F1 Surgical Ward
This is a 56 year old woman who had a laparotomy yesterday for resection of her sigmoid colon. She was previously fit and well, the procedure was complicated by faecal soiling of the peritoneum. She has been increasingly short of breath overnight and is now restless and confused. Here is her observation chart. She has ECG, a pulse oximeter and non-invasive blood pressure monitoring attached.

(Sepsis and Acute Respiratory Distress Syndrome (ARDS) on CXR if requested)

Scenario 4

F1 Surgical Ward
This 32 year old woman had a laparotomy 2 days ago for a perforated appendix. At the operation free pus was found in the peritoneum but she made an uneventful recovery. This morning about 30 minutes after an intravenous dose of antibiotics she has become short of breath, wheezy, flushed and feels faint on sitting up. She has ECG, a pulse oximeter and non-invasive blood pressure monitoring attached.

(? late anaphylaxis, ? anastomotic leak, gas under diaphragm on CXR if requested)

Scenario 5

F1 Surgical Ward
This is a 75 year old man with a history of a previous myocardial infarct and treated hypertension. He had an inguinal hernia repaired yesterday under general anaesthesia. He recovered well from the anaesthesia and returned to the surgical ward post-operatively. He has not passed urine since the operation. He feels slightly short of breath. He has ECG, a pulse oximeter and non-invasive blood pressure monitoring attached

(Acute renal failure, rising potassium and metabolic acidosis if requested)

Scenario 6
F1 Emergency Department
This young woman has been brought in by her mother as she has been feeling unwell on and off for the last 4 days. She has a pulse oximeter and non-invasive blood pressure attached.

(Supra-ventricular tachycardia, adenosine, verapamil, DC cardioversion if requested)

Scenario 7 (Final assessment)

F1 Emergency Department
This elderly gentleman has been brought in by the paramedics. He was apparently well yesterday but was found unconscious in his bathroom this morning. He has a pulse oximeter and non-invasive blood pressure attached.

(Hypothermia, 34C, sinus bradycardia, normal blood sugar, causes of loss of consciousness)

APPENDIX 5.2. THE MOLTI STUDY BASELINE QUESTIONNAIRE

UNIVERSITY OF BIRMINGHAM

The Management of Life Threatening Illness (MOLTI) course: An efficacy assessment on the SimMan

Baseline Assessment Participant Questionnaire

We would be very grateful if you could answer the following questions on your demographics and previous experience. All information collected will remain confidential and no links to participants will be made**. All data will be held securely in line with the Data Protection Act**. PLEASE PRINT.

12. Name: ……...
13. Sex (Please tick): Male Female
14. Date of Birth ……/……/……
15. City/Country of Birth……..
16. Ethnicity:..……...
17. Have you attended a MOLTI course in the past?

    Yes No

If yes, in which month of your Year 5 studies?.....

and in which hospital ?.....

18. Have you completed your ILS course?

    Yes No

If yes, in which month of your Year 5 studies?.....

and in which hospital ?.....

19. Have you attended an ALERT course in the past?

    Yes No

If yes, in which month of your Year 5 studies?.....

and in which hospital?.....

20. Have you had previous training with a computer-based simulator?

    Yes No

If yes, please describe the occasion(s) in terms of

- year of your studies:
- hospital:
- type of simulator:
- content of the training:

10. If I was called to see a deteriorating patient, I am confident I could start the right basic management (Please circle the appropriate).

Strongly disagree	**Disagree**	**Undecided**	**Agree**	**Strongly agree**
1	**2**	**3**	**4**	**5**

APPENDIX 5.3. SAMPLE OF THE MOLTI STUDY BASELINE ASSESSMENT SIMMAN SOFTWARE OUTPUT

00:00:00 Name: Sally Age: 28 years Weight: 65 kg Height: 170 cm Gender: Female Description: Known mild asthma for 10 years, no previous hospital admissions Increasing shortness of breath for last 12 hours Now feels dreadful and unable to talk She has and ECG, a pulse oximeter and non-invasive blood pressure monitoring attached

00:00:00 Trend started: MOLTI Level 1 Trend 1
00:00:00 HR: 80 BP = 120/ 80 SpO2: 98
00:00:00 Simulation paused
00:00:00 Simulation resumed
00:00:00 Vocal sound = SOBBreathing
00:00:09 Heart rate = 81
00:00:13 Respiration rate = 14
00:00:25 Vocal sound = Difficult breathing
00:00:31 BP = 110/ 70
00:00:43 SpO2 = 91
00:00:44 Alarm: SpO2 LOW
00:00:57 Oxygen
00:00:57 Frame: SO2 Recovery
00:00:57 Trend started: MOLTI Level 1SO2 Recovery
00:01:00 Heart rate = 97
00:01:00 HR: 97 BP = 100/ 60 SpO2: 91
00:01:13 SpO2 = 97
00:01:19 Respiration rate = 24
00:01:28 NBP measured. BP : 91/51
00:01:31 Monitor alarms silenced
00:01:37 BP = 88/ 48
00:01:46 Heart rate = 114
00:02:00 HR: 119 BP = 80/ 40 SpO2: 100
00:02:04 Alarm: NBPs LOW
00:02:04 NBP measured. BP : 79/39
00:02:29 Heart rate = 131
00:02:29 Alarm: HR HIGH
00:02:30 IV Inserted
00:02:31 Respiration rate = 35
00:02:31 BP = 70/ 30
00:02:37 Volume Infusion
00:02:37 Frame: BP Recovery
00:02:37 Trend started: BP recovery
00:02:53 Vocal sound = Difficult breathing
00:03:00 HR: 132 BP = 66/ 26 SpO2: 100
00:03:23 Salbutamol Neb
00:03:24 BP = 103/ 66
00:03:37 Respiration rate = 45
00:03:55 Heart rate = 116
00:04:00 HR: 118 BP = 102/ 65 SpO2: 100
00:04:08 NBP measured. BP : 101/64
00:04:08 Order chest x-ray

00:04:08 Frame: Tension Pneumothorax
00:04:08 Vocal sound = I feel really bad
00:04:08 Trend stopped: MOLTI Level 1 Trend 1
00:04:08 Trend stopped: MOLTI Level 1SO2 Recovery
00:04:08 Trend stopped: BP recovery
00:04:08 Trend started: Tension Pneumothorax
00:04:08 X-ray presented: Scenario 1 CXR 1.jpg
00:04:19 BP = 83/ 46
00:04:21 Heart rate = 132
00:04:22 SpO2 = 92
00:04:25 BP = 70/ 33
00:04:25 Respiration rate = 57
00:04:25 Alarm: SpO2 LOW
00:04:31 Alarm: EXTREME TACHY
00:04:31 BP = 58/ 21
00:04:33 Heart rate = 153
00:04:34 SpO2 = 84
00:04:37 BP = 45/ 8
00:04:45 Heart rate = 171
00:04:46 SpO2 = 76
00:04:49 BP = 20/ 0
00:04:58 Frame: Recovery
00:04:58 Trend stopped: Tension Pneumothorax
00:04:58 Trend started: MOLTI Recover All
00:04:58 Heart rate = 193
00:04:58 SpO2 = 68
00:05:00 HR: 194 BP = 2/ 0 SpO2: 68
00:05:04 Pneumothorax Decompression
00:05:06 BP = 12/ 10
00:05:06 SpO2 = 77
00:05:12 BP = 22/ 20
00:05:12 SpO2 = 87
00:05:18 BP = 32/ 30
00:05:24 BP = 42/ 41
00:05:25 SpO2 = 100
00:05:30 BP = 52/ 51
00:05:31 Heart rate = 175
00:05:35 BP = 135/ 61
00:05:42 BP = 145/ 88
00:05:42 Respiration rate = 45
00:05:48 Heart rate = 123
00:05:48 BP = 155/ 98
00:05:54 Heart rate = 107
00:05:54 BP = 165/ 108
00:06:00 NBP measured. BP : 176/119
00:06:00 BP = 176/ 119
00:06:00 HR: 105 BP = 176/ 119 SpO2: 100
00:06:01 Simulation ended - Go to Debrief

APPENDIX 5.4. THE MOLTI STUDY BASELINE ASSESSMENT CHECKLIST

The MOLTI study: Sally (Asthma, Tension Pneu) **Student ID:…….**

Item	Marks available	Marks awarded
Assessed Airway and Breathing	2	
Assessed Circulation	2	
Gave O_2 (no marks if prompted)	2	
Used a non-rebreather (or Hudson) mask to give O_2	1	
Gave high flow O_2 (10-15 lpm)	2	
Gave O_2 prior to the SpO2 'low' alarm	2	
Asked for BP measurement (no marks if prompted)	2	
Asked for IV access (no marks if prompted)	2	
Gave IV fluids prior to the NBPs 'low' alarm	1	
Gave appropriate type of IV fluids (NOT 5% Dextrose)	1	
Gave appropriate volume of IV fluids (aprox 500 ml)	1	
Asked for help	1	
Gave Nebulised salbutamol	1	
Correctly diagnosed tension pneumothorax	2	
Asked for appropriate venflon (i.e. orange or brown or 14G)	2	
Indicated correctly where venflon should go	2	
	Negative marks	**Marks awarded**
Adopted a 'shotgun' approach (i.e asked for many investigations/interventions all at once)	- 1	
Asked for inappropriate intervention e.g. urinary catheter	- 1	

APPENDIX 5.5. THE MOLTI STUDY GLOBAL RATING TOOL

Generic Management Global Rating Scale

Participant ID:……..

Assessor's Name :……

Please score the participant's performance (please circle the appropriate):

	poor	borderline	fair	good	excellent
Medical Management	**1**	**2**	**3**	**4**	**5**
Gathers the appropriate information Institutes appropriate interventions Interventions are timely Appropriate order of interventions Reaches diagnosis, considers differential					
Behaviour	**1**	**2**	**3**	**4**	**5**
Acquisition of all available information Anticipates and plans Calls for help appropriately Re-evaluates situation Utilizes resources effectively Allocates attention wisely (i.e. does not get distracted away from ABC-O_2, I.V.-D) Prioritises Concise, directed instructions Does not adopt a 'shotgun'approach					
Overall performance	**1**	**2**	**3**	**4**	**5**

Overall would you consider this student's performance as a Pass or a Fail? (Please tick)

Pass ☐ **Fail** ☐

APPENDIX 5.6. SAMPLE OF THE MOLTI STUDY FINAL ASSESSMENT SIMMAN SOFTWARE OUTPUT

00:00:00 Name: Denis Age: 85 years Weight: 86 kg Height: 180 cm Gender: Male
Description: This elderly gentleman has been brought in by the paramedics. He was apparently well yesterday but was found unconscious in his bathroom this morning. He has a pulse oximeter and non-invasive blood pressure attached

00:00:00 Trend started: MOLTI Level 1 Trend 1
00:00:00 HR: 40 BP = 90/ 40 SpO2: 85
00:00:00 Simulation paused
00:00:00 Simulation resumed
00:00:03 Alarm: SpO2 LOW
00:00:04 Alarm: HR LOW
00:00:09 Heart rate = 41
00:00:12 Respiration rate = 14
00:00:19 Check airway patency
00:00:20 Check breathing
00:00:22 Oxygen
00:00:22 Frame: SO2 Recovery
00:00:22 Trend started: MOLTI Level 1SO2 Recovery
00:00:30 BP = 80/ 30
00:00:34 Trend stopped: MOLTI Level 1 Trend 1
00:00:38 Heart rate = 78
00:00:42 SpO2 = 92
00:00:50 NBP measured. BP : 78/28
00:00:51 Heart rate = 52
00:00:51 Alarm: NBPs LOW
00:00:54 SpO2 = 100
00:01:00 HR: 65 BP = 78/ 28 SpO2: 100
00:01:05 Heart rate = 75
00:01:06 Measure glucose
00:01:11 IV Inserted
00:01:11 Volume Infusion
00:01:12 Heart rate = 58
00:01:37 rapid infusion
00:01:41 Monitor alarms silenced
00:01:48 Order chest x-ray
00:01:48 Frame: CXR
00:02:00 HR: 76 BP = 78/ 28 SpO2: 100
00:02:13 Blood cultures, LFT, FBC,
00:02:25 NBP measured. BP : 78/28
00:02:28 Temp 33.5
00:02:51 BP = 104/ 69
00:02:54 Heart rate = 88
00:03:00 Bear Hugger
00:03:00 HR: 70 BP = 104/ 69 SpO2: 100
00:03:47 Warm fluids
00:04:00 HR: 74 BP = 104/ 69 SpO2: 100
00:04:12 NBP measured. BP : 104/69
00:04:13 Request blood

00:04:20 4 blood
00:04:32 Heart rate = 53
00:04:49 Measure glucose
00:04:54 Heart rate = 74
00:04:58 Carotid pulse check. Pulse strength: normal
00:05:00 HR: 75 BP = 104/ 69 SpO2: 100
00:05:07 Heart rate = 51
00:05:12 Simulation paused
00:05:12 Simulation ended - Go to Debrief

APPENDIX 5.7. THE MOLTI STUDY FINAL ASESSMENT CHECKLIST

The MOLTI study: Dennis (Hypothermia) **Student ID:…….**

Item	Marks available	Marks awarded
Assessed Airway and Breathing	2	
Assessed Circulation	2	
Gave O_2 (no marks if prompted)	2	
Used a non-rebreather (or Hudson) mask to give O_2	1	
Gave high flow O_2 (10-15 lpm)	2	
Asked for BP measurement (no marks if prompted)	2	
Asked for IV access (no marks if prompted)	2	
Gave IV fluids prior to the NBPs 'low' alarm	1	
Gave appropriate type of IV fluids (NOT 5% Dextrose)	1	
Gave appropriate volume of IV fluids (500ml STAT)	1	
Asked for blood glucose measurement (no marks if prompted)	2	
Used the Glasgow Coma Scale	1	
Asked for FBC	1	
Asked for U&Es	1	
Asked for help	1	
Asked for blood tests to be done while putting I.V.	1	
Checked for bruises/trauma	1	
Asked for alcohol level /drug screen	1	
Asked for temperature measurement (no marks if prompted)	1	
Asked for blankets, warming devices, warm IV fluids	2	
	Negative marks	**Marks awarded**
Adopted a 'shotgun' approach (i.e asked for many investigations/interventions all at once)	-1	
Asked for inappropriate intervention e.g. urinary catheter	- 1	

APPENDIX 5.8. THE MOLTI STUDY FEEDBACK QUESTIONNAIRE

UNIVERSITY OF BIRMINGHAM

The Management of Life Threatening Illness (MOLTI) course: An efficacy assessment on the SimMan

Participant Feedback Questionnaire

Name:……

We would like your feedback with regards to the following statements. Please circle the appropriate:

Group:…… *Intervention/ Control*……

1st assessment	**Strongly disagree**	**Disagree**	**Undecided**	**Agree**	**Strongly agree**
1. I found the orientation to the simulator adequate	1	2	3	4	5
2. The information provided by the instructor was clear during the scenario	1	2	3	4	5
3. My stock of knowledge was adequate to the task	1	2	3	4	5
4. I was able to manage the scenario based on reading /general knowledge without prior experience	1	2	3	4	5
5. Overall I feel that my baseline assessment would be rated as	Poor	Borderline	Fair	Good	Excellent
The MOLTI course	**Strongly disagree**	**Disagree**	**Undecided**	**Agree**	**Strongly agree**
6. What I learned from the scenarios was more than the knowledge I brought to it	1	2	3	4	5
7. The simulator enhanced learning more than reading would	1	2	3	4	5
8. I felt I did things I would never have a chance to practice otherwise	1	2	3	4	5
9. The debriefing provided logically organised feedback and clarified important issues of the scenario	1	2	3	4	5
10. The debriefing provided enhanced my stock of knowledge	1	2	3	4	5
11. I expect that the knowledge gained from the scenarios will be helpful to me in practice	1	2	3	4	5
12. If I was called to see a deteriorating patient, I am confident I could start the right basic management	1	2	3	4	5

2nd assessment	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
13. The information provided by the instructor was clear during the scenario	1	2	3	4	5
14. My stock of knowledge was adequate to the task	1	2	3	4	5
15. Overall I feel that my 2nd assessment would be rated as	Poor	Borderline	Fair	Good	Excellent

Overall	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
16. If I was called to a deteriorating patient I am worried I would do the wrong thing	1	2	3	4	5
17. I felt comfortable with the simulator environment	1	2	3	4	5
18. I felt that the simulation environment and scenarios prompted realistic responses from me	1	2	3	4	5
19. I found it easy to treat the mannequin as a simulated human	1	2	3	4	5
20. The presence of the scenario director detracted from the realism of the simulator experience	1	2	3	4	5
21. The video camera interfered with the simulator experience	1	2	3	4	5
22. Overall, I was satisfied with the use of the human patient simulator for training	1	2	3	4	5
23. I found the MOLTI course useful	1	2	3	4	5
24. I enjoyed the MOLTI course	1	2	3	4	5

We would also be very grateful for your comments regarding the following:

25. What did you like/find helpful about the MOLTI course?

……

……

……

……

……

26. What would you like to suggest to make the MOLTI course better?

……

……

……

……

……

27. Additional comments:

……

……

……

……

……

28. Should these sessions be continued for future students?

……

……

THANK YOU VERY MUCH FOR YOUR HELP!

BIBLIOBRAPHY

1. Torkington J, Smith SG, Rees BI, Darzi A. The role of simulation in surgical training. Annals of the Royal College of Surgeons of England 2000; 82(2):88-94

2. Maran NJ, Glavin RJ. Low- to high-fidelity simulation - A continuum of medical education? Medical Education 2003; 37(1):22-28

3. DeVita MA. Simulation and the prognosis for the apprenticeship model of health care education. Journal of Intensive Care Medicine 2007; 22(5):310-311

4. Steketee C, Bower A. Augmenting the clinical apprenticeship model. Proceedings of AARE 2007; 1-13

5. Walter AJ. Surgical education for the twenty-first century: beyond the apprentice model. Obstetrics and Gynecology Clinics of North America 2006; 33(2):233-236

6. Kneebone R. Simulation in surgical training: educational issues and practical implications. Medical Education 2003; 37(3):267-277

7. Dent JA. A practical guide for medical teachers. 1st ed. London: Harcourt Publishers Limited; 2001

8. Tsang JS, Naughton PA, Leong S, Hill DK, Kelly CJ, Leahy AL. Virtual reality simulation in endovascular surgical training. Surgeon Journal of the Royal Colleges of Surgeons of Edinburgh & Ireland 2008; 6(4):214-220

9. Vozenilek J, Huff JS, Reznek M, Gordon JA. See one, do one, teach one: advanced technology in medical education. Academic Emergency Medicine 2004; 11(11):1149-1154

10. Hammond J. Simulation in critical care and trauma education and training. Current Opinion in Critical Care 2004; 10(5):325-329

11. Kneebone R. Evaluating clinical simulations for learning procedural skills: a theory-based approach. Academic Medicine 2005; 80(6):549-553

12. Department of Health. The European working time directive for trainee doctors. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_093941.pdf. 2009

13. Mathers J, Parry J, Lewis S, Greenfield S. What impact will the 'conversion' of two district general hospitals into teaching hospitals have? Views from the field. Medical Education 2003; 37(3):223-232

14. Bradley P. The history of simulation in medical education and possible future directions. Medical Education 2006; 40(3):254-262

15. McGaghie WC, Siddall VJ, Maxmanian PE, Myers J. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education. Chest 2009; 135(Suppl 3):62S-68S

16. Issenberg SB, McGaghie WC, Petrusa ER, Gordon DL, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Medical Teacher 2005; 27(1):10-28

17. Adrales GL, Park AE, Chu UB et al. A valid method of laparoscopic simulation training and competence assessment. Journal of Surgical Research 2003; 114(2):156-162

18. Department of Health. A high quality workforce: NHS next stage review. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_085841.pdf. 2008

19. General Medical Council. Tomorrow's doctors. http://www.gmc-uk.org/static/documents/content/TomorrowsDoctors_2009.pdf. 2009

20. Streufert S, Satish U, Barach P. Improving medical care: the use of simulation technology. Simulation & Gaming 2001; 32(2):164-174

21. Gerson LB. Evidence-based assessment of endoscopic simulators for training. Gastrointestinal Endoscopy Clinics of North America 2006; 16(3):489-509

22. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. Academic Medicine 2003; 78(8):783-788

23. Fletcher GCL, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. British Journal of Anaesthesia 2002; 88(3):418-429

24. Salas E, Cannon-Bowers JA. The science of training: a decade of progress. Annual Review of Psychology 2001; 52:471-499

25. Regehr G. Trends in medical education research. Academic Medicine 2004; 79(10):939-947

26. DeVita MA, Schaefer J, Lutz J, Wang H, Dongilli T. Improving medical emergency team (MET) performance using a novel curriculum and a computerized human patient simulator. Quality and Safety in Health Care 2005; 14(5):326-331

27. Alinier G. A typology of educationally focused medical simulation tools. Medical Teacher 2007; 29(8):e243-e250

28. Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? Quality and Safety in Health Care 2004; 13(Suppl 1):i51-i56

29. Cumin D, Merry AF. Simulators for use in anaesthesia. Anaesthesia 2007; 62(2):151-162

30. Meller G. A typology of simulators for medical education. Journal of Digital Imaging 1997(3 Suppl 1); 10:194-196

31. Lighthall GK, Barr J. The use of clinical simulation systems to train critical care physicians. Journal of Intensive Care Medicine 2007; 22(5):257-269

32. Roberts KE, Bell RL, Duffy AJ. Evolution of surgical skills training. World Journal of Gastroenterology 2006; 12(20):3219-3224

33. Beullens J, Rethans JJ, Goedhuys J, Buntinx F. The use of standardized patients in research in general practice. Family Practice 1997; 14(1):58-62

34. Hamstra SJ, Dubrowski A, Backstein D. Teaching technical skills to surgical residents. Clinical Orthopaedics & Related Research 2006; 449:108-115

35. Seropian MA. General concepts in full scale simulation: getting started. Anesthesia & Analgesia 2003; 97(6):1695-1705

36. Jha AK, Duncan BW, Bates DW. Chapter 45. Simulator-based training and patient safety. In: Shojania KG, Duncan BW, McDonald KM, et al, editors. Making Health Care Safer: A Critical Analysis of Patient Safety Practices. Evidence Report/Technology Assessment No. 43 (Prepared by the University of California at San Francisco-Stanford Evidence-based Practice Center under Contract No. 290-97-0013), AHRQ Publication No. 01-E058, Rockville, MD: Agency for Healthcare Research and Quality. Rockville, MD: Agency for Healthcare Research and Quality; 2001

37. McFetrich J. A structured literature review on the use of high fidelity patient simulators for teaching in emergency medicine. Emergency Medicine Journal 2006; 23(7):509-511

38. Issenberg SB, McGaghie WC, Hart IR et al. Simulation technology for health care professional skills training and assessment. JAMA 1999; 282(9):861-866

39. Brown C, Hofer T, Johal A et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 1. Conceptualising and developing interventions. Quality and Safety in Health Care 2008; 17(3):158-162

40. Goodwin MWP, French GWG. Simulation as a training and assessment tool in the management of failed intubation in obstetrics. International Journal of Obstetric Anesthesia 2001; 10(4):273-277

41. Moorthy K, Smith S, Brown T, Bann S, Darzi A. Evaluation of virtual reality bronchoscopy as a learning and assessment tool. Respiration 2003; 70(2):195-199

42. Kneebone R, Nestel D, Vincent C, Darzi A. Complexity, risk and simulation in learning procedural skills. Medical Education 2007; 41(8):808-814

43. Ostergaard D. National medical simulation training program in Denmark. Critical Care Medicine 2004; 32(Suppl 2):S58-S69

44. Kolb DA, Boyatzis RE, Mainemelis C. Experiential learning theory: previous research and new directions. In: Sternberg RJ, Zhang LF, editors. Perspectives on cognitive, learning, and thinking styles. NJ: Lawrence Erlbaum; 2000

45. Bradley P, Postlethwaite K. Simulation in clinical learning. Medical Education 2003; 37(Suppl 1):1-5

46. Perkins GD. Simulation in resuscitation training. Resuscitation 2007; 73(2):202-211

47. Issenberg SB, Scalese RJ. Best evidence on high-fidelity simulation: what clinical teachers need to know. The Clinical Teacher 2007; 4(2):73-77

48. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic Medicine 2004; 79(Suppl 10):S70-S81

49. Dent JA, Harden RM. A practical guide for medical teachers. Edinburgh: Elsevier; 2005

50. Norman G, Schmidt HG. Effectiveness of problem-based learning curricula: theory, practice and paper darts. Medical Education 2000; 34(9):721-728

51. Phillips R. Challenging the primacy of lectures: the dissonance between theory and practice in university teaching. Journal of University Teaching and Learning Practice 2 (1). http://jutlp.uow.edu.au/2005_v02_i01/phillips003.html. 2005

52. Gibbs G. Improving the quality of student learning. Bristol: Technical and Educational Services Lts; 1992

53. McDougall EM. Validation of Surgical Simulators. Journal of Endourology 2007; 21(3):244-247

54. Brown C, Hofer T, Johal A et al. An epistemology of patient safety research: a framework for study design and interpretation. Part 2. Study design. Quality and Safety in Health Care 2008; 17(3):164-171

55. Torgerson CJ. Educational research and randomised trials. Medical Education 2002; 36(11):1002-1003

56. Lammers RL, Davenport M, Korley F et al. Teaching and assessing procedural skills using simulation: metrics and methodology. Academic Emergency Medicine 2008; 15(11):1079-1087

57. Foray D, Murnane R, Nelson R. Randomized trials of education and medical practices: strengths and limitations. Economics of Innovation and New Technology 2007; 16(5):303-306

58. Bligh J, Brice J. What is the value of good medical education research? Medical Education 2008; 42(7):652-653

59. Gill D, Griffin AE. Reframing medical education research: let's make the publishable meaningful and the meaningful publishable. Medical Education 2009; 43(10):933-935

60. Greenhalgh T, Taylor R. How to read a paper. Papers that go beyond numbers (qualitative research). BMJ 1997; 315(7110):740-743

61. Glavin RJ, Maran NJ. Editorial I: Development and use of scoring systems for assessment of clinical competence. British Journal of Anaesthesia 2002; 88(3):329-330

62. Miller GE. The assessment of clinical skills/competence/performance. Academic Medicine 1990; 65(Suppl 9):S63-S67

63. Tavakol M, Mohagheghi MA, Dennick R. Assessing the skills of surgical residents using simulation. Journal of Surgical Education 2008; 65(2):77-83

64. Byrne AJ, Greaves JD. Assessment instruments used during anaesthetic simulation: review of published studies. British Journal of Anaesthesia 2001; 86(3):445-450

65. McAleer S. Choosing assessment instruments. In: Dent JA, Harden RM, editors. A practical guide for medical teachers. 2nd ed. Edinburgh: Elsevier; 2005 p. 302-310

66. Schout BMA, Hendrikx AJM, Scherpbier AJJA, Bemelmans BLH. Update on Training Models in Endourology: A Qualitative Systematic Review of the Literature between January 1980 and April 2008. European Urology 2008; 54(6):1247-1261

67. Wong AK. Full scale computer simulators in anesthesia training and evaluation. Canadian Journal of Anaesthesia 2004; 51(5):455-464

68. Rousson V, Gasser T, Burkhardt S. Assessing intrarater, interrater and test-retest reliability of continuous measurements. Statistics in Medicine 2002; 21(22):3431-3446

69. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Statistical assessment of the learning curves of health technologies. Health Technology Assessment 2001; 5(12):1-2

70. Ellis S, Morgan A, Taylor L, Walters R, Goodrich J. HIV prevention: a review of reviews assessing the effectiveness of interventions to reduce the sexual risk of HIV transmission. 1st ed. London: Health Development Agency; 2003

71. Haque S, Srinivasan S. A meta-analysis of the training effectiveness of virtual reality surgical simulators. IEEE Transactions of Information Technology in Biomedicine 2006; 10(1):51-58

72. Barach P, Bentham L, Bion J et al. Teaching life support and resuscitation competencies in health care - current practice and strategies for future research. LIST report. Birmingham. 2005

73. Arnold P, Farrell MJ. Can virtual reality be used to measure and train surgical skills? Ergonomics 2002; 45(5):362-379

74. Aucar JA, Groch NR, Troxel SA, Eubanks SW. A review of surgical simulation with attention to validation methodology. Surgical Laparoscopy, Endoscopy & Percutaneous Techniques 2005; 15(2):82-89

75. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. British Journal of Surgery 2008; 95(9):1088-1097

76. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ. A systematic review of skills transfer after surgical simulation training. Annals of Surgery 2008; 248(2):166-179

77. Sutherland LM, Middleton PF, Anthony A et al. Surgical simulation: a systematic review. Annals of Surgery 2006; 243(3):291-300.

78. Gaffan J, Dacre J, Jones A. Educating undergraduate medical students about oncology: a literature review. Journal of Clinical Oncology 2006; 24(12):1932-1939

79. Hamilton R. Nurses' knowledge and skill retention following cardiopulmonary resuscitation training: a review of the literature. Journal of Advanced Nursing 2005; 51(3):288-297

80. Lane C, Rollnick S. The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. Patient Education and Counseling 2007; 67(1-2):13-20

81. Ravert P. An integrative review of computer-based simulation in the education process. CIN: Computers, Informatics, Nursing 2002; 20(5):203-208

82. Byrne AJ, Pugsley L, Hashem MA. Review of comparative studies of clinical skills training. Medical Teacher 2008; 30(8):764-767

83. Lynagh M, Burton R, Sanson-Fisher R. A systematic review of medical skills laboratory training: Where to from here? Medical Education 2007; 41(9):879-887

84. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. Effect of practice on standardised learning outcomes in simulation-based medical education. Medical Education 2006; 40(8):792-797

85. Matthes K. Simulator Training in Endoscopic Hemostasis. Gastrointestinal Endoscopy Clinics of North America 2006; 16(3):511-527

86. Kefalides PT, Gress F. Simulator Training for Endoscopic Ultrasound. Gastrointestinal Endoscopy Clinics of North America 2006; 16(3):543-552

87. Baillie J. Endoscopic Retrograde Cholangiopancreatography Simulation. Gastrointestinal Endoscopy Clinics of North America 2006; 16(3):529-542

88. Gallagher AG, Ritter EM, Champion H et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. Annals of Surgery 2005; 241(2):364-372

89. Hammond I, Karthigasu K. Training, assessment and competency in gynaecologic surgery. Best Practice & Research in Clinical Obstetrics & Gynaecology 2006; 20(1):173-187

90. Reznek M, Harter P, Krummel T. Virtual reality and simulation: training the future emergency physician. Academic Emergency Medicine 2002; 9(1):78-87

91. Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. British Journal of Surgery 2004; 91(12):1549-1558

92. Lake CL. Simulation in cardiothoracic and vascular anesthesia education: tool or toy? Seminars in Cardiothoracic & Vascular Anesthesia 2005; 9(4):265-273

93. Mantovani F, Castelnuovo G, Gaggioli A, Riva G. Virtual reality training for health-care professionals. Cyberpsychology & Behavior 2003; 6(4):389-395

94. Khalifa YM, Bogorad D, Gibson V, Peifer J, Nussbaum J. Virtual Reality in Ophthalmology Training. Survey of Ophthalmology 2006; 51(3):259-273

95. Laguna MP, Hatzinger M, Rassweiler J. Simulators and endourological training. Current Opinion in Urology 2002; 12(3):209-215

96. Dierdorf SF. Education in the use of the laryngeal mask airway. International Anesthesiology Clinics 1998; 36(2):19-29

97. du Boulay C, Medway C. The clinical skills resource: a review of current practice. Medical Education 1999; 33(3):185-191

98. Kneebone R, Nestel D, Wetzel C et al. The human face of simulation: Patient-focused simulation training. Academic Medicine 2006; 81(10):919-924

99. Ander DS, Hanson A, Pitts S. Assessing resident skills in the use of rescue airway devices. Annals of Emergency Medicine 2004; 44(4):314-319

100. Tiah L, Wong E, Chen MFJ, Sadarangani SP. Should there be a change in the teaching of airway management in the medical school curriculum? Resuscitation 2005; 64(1):87-91

101. Vertongen VM, Ramsay MP, Herbison P. Skills retention for insertion of the Combitube and laryngeal mask airway. Emergency Medicine 2003; 15(5-6):459-464

102. Weksler N, Tarnopolski A, Klein M et al. Insertion of the endotracheal tube, laryngeal mask airway and oesophageal-tracheal Combitube. A 6-month comparative prospective study of acquisition and retention skills by medical students. European journal of Anaesthesiology 2005; 22(5):337-340

103. Davies PR, Tighe SQ, Greenslade GL, Evans GH. Laryngeal mask airway and tracheal tube insertion by unskilled personnel. Lancet 1990; 336(8721):977-979

104. Murray JM, Vermeulen MJ, Morrison LJ, Waite T. Evaluation of prehospital insertion of the laryngeal mask airway by primary care paramedics with only classroom mannequin training. Canadian Journal of Emergency Medicine 2002; 4(5):338-343

105. Frascone RJ, Pippert G, Heegaard W, Molinari P, Dries D. Successful training of HEMS personnel in laryngeal mask airway and intubating laryngeal mask airway placement. Air medical journal 2008; 27(4):185-187

106. Morse Z, Sano K, Kageyama I, Kanri T. The relationship of placement accuracy and insertion times for the laryngeal mask airway to the training of inexperienced dental students. Anesthesia Progress 2002; 49(1):9-13

107. Altman DG, Bland JM. How to randomise. BMJ 1999; 319(7211):703-704

108. The Laryngeal Mask Company Limited. LMA airway instruction manual. San Diego, CA. 2005

109. Altman DG. Practical statistics for medical research. 1st ed. Cornwall: Chapman and Hall; 1991

110. Hamaguchy S, Egawa H, Okuda Y, Kitajima T. Use of a fiberscope and closed-circuit television for teaching laryngeal mask insertion. Anesthesia & Analgesia 2000; 91(2):501

111. Roberts I, Allsop P, Dickinson M, Curry P, Eastwick-Field P, Eyre G. Airway management training using the laryngeal mask airway: a comparison of two different training programmes. Resuscitation 1997; 33(3):211-214

112. Durmaz E. Theatre-based hospital placements for firefighters. British Journal of Perioperative Nursing 2005; 15(9):396-399

113. Parry K, Owen H. Small simulators for teaching procedural skills in a difficult airway algorithm. Anaesthesia & Intensive Care 2004; 32(3):401-409

114. Stringer KR, Bajenov S, Yentis SM. Training in airway management. Anaesthesia 2002; 57(10):967-983

115. Smith GB, Osgood VM, Crane S. ALERT^{TM} - a multiprofessional training course in the care of the acutely ill adult patient. Resuscitation 2002; 52(3):281-286

116. McQuillan PJ, Pilkington S, Allan A et al. Confidential inquiry into quality of care before admission to intensive care. BMJ 1998; 316(7148):1853-1858

117. Hillman K, Chen J, Cretikos M et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. The Lancet 2005; 365(9477):2091-2097

118. Priestley G, Watson W, Rashidian A et al. Introducing Critical Care Outreach: a ward-randomised trial of phased introduction in a general hospital. Intensive Care Medicine 2004; 30(7):1398-1404

119. Murray D. Clinical skills in acute care: A role for simulation training. Critical Care Medicine 2006; 34(1):252-253

120. Weller J, Robinson B, Larsen P, Caldwell C. Simulation-based training to improve acute care skills in medical undergraduates. The New Zealand Medical Journal 117(1204). http://www.nzma.org.nz/journal/117-1204/1119/. 2004

121. Alinier G, Hunt B, Gordon R, Harwood C. Effectiveness of intermediate-fidelity simulation training technology in undergraduate nursing education. Journal of Advanced Nursing 2006; 54(3):359-369

122. Steadman RH, Coates WC, Huang YM et al. Simulation-based training is superior to problem-based learning for the acquisition of critical assessment and management skills. Critical Care Medicine 2006; 34(1):151-157

123. Morgan PJ, Cleave-Hogg D, Mcllroy J, Devitt JH. A comparison of experiential and visual learning for undergraduate medical students. Anesthesiology 2002; 96(1):10-6

124. MacDowall J. The assessment and treatment of the acutely ill patient - the role of the patient simulator as a teaching tool in the undergraduate programme. Medical Teacher 2006; 28(4):326-329

125. Weller J. Simulation in undergraduate medical education: bridging the gap between theory and practice. Medical Education 2004; 38(1):32-38

126. Morgan PJ, Cleave-Hogg D. Evaluation of medical students' performance using the anaesthesia simulator. Medical Education 2000; 34(1):42-45

127. Bearnson CS, Wiker KM. Human patient simulators: A new face in baccalaureate nursing education at Bringham Young University. Journal of Nursing Education 2005; 44(9):421-425

128. Weller JM, Bloch M, Young S et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. British Journal of Anaesthesia 2003; 90(1):43-47

129. Holzman RS, Cooper JB, Gaba DM, Philip JH, Small SD, Feinstein D. Anesthesia crisis resource management: real-life simulation training in operating room crises. Journal of Clinical Anesthesia 1995; 7(8):675-687

130. Boulet JR, Murray D, Kras J, Woodhouse J, McAllister J, Ziv A. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. Anesthesiology 2003; 99(6):1270-1280

131. Cherry RA, Williams J, George J, Ali J. The Effectiveness of a Human Patient Simulator in the ATLS Shock Skills Station. Journal of Surgical Research 2007; 139(2):229-235

132. Morgan PJ, Cleave-Hogg D. Comparison between medical students' experience, confidence and competence. Medical Education 2002; 36(6):534-539.

133. Cook DA. Avoiding confounded comparisons in education research. Medical Education 2009; 43(2):102-104

134. The Higher Education Academy. The UK Professional Standards Framework for teaching and supporting learning in higher education. http://www.heacademy.ac.uk/assets/York/documents/ourwork/professional/ProfessionalStandardsFramework.pdf. 2006

135. Michelson JD, Manning L. Competency assessment in simulation-based procedural education. The American Journal of Surgery 2008; 196(4):609-615

136. Anderson JM, Aylor ME, Leonard DT. Instructional design dogma: creating planned learning experiences in simulation. Journal of Critical Care 2008; 23(4):595-602

137. Norman G. The american college of chest physicians evidence-based educational guidelines for continuing medical education interventions a critical review of evidence-based educational guidelines. Chest 2009; 135(3):834-837

138. Norman G, Dore K, Krebs J, Neville AJ. The power of the plural: effect of conceptual analogies on successful transfer. Academic Medicine 2007; 82(10):S16-S18

139. Eppich WJ, Adler MD, McGaghie WC. Emergency and critical care pediatrics: use of medical simulation for training in acute pediatric emergencies. Current Opinion in Pediatrics 2006; 18(3):266-271

140. Fitzgerald TN, Duffy AJ, Bell RL, Berman L, Longo WE, Roberts KE. Computer-based endoscopy simulation: emerging roles in teaching and professional skills assessment. Journal of Surgical Education 2008; 65(3):229-235

141. Goldmann K, Ferson DZ. Education and training in airway management. Best Practice & Research 2005; 19(4):717-732

142. Gould DA, Kessel DO, Healey AE, Johnson SJ, Lewandowski WE. Simulators in catheter-based interventional radiology: training or computer games? Clinical Radiology 2006; 61(7):556-561

143. Hart R, Karthigasu K. The benefits of virtual reality simulator training for laparoscopic surgery. Current Opinion in Obstetric Gynecology 2007; 19(4):297-302

144. Laguna MP, de Reijke TM, Wijkstra H, de la Rosette J. Training in laparoscopic urology. Current Opinion in Urology 2006; 16(2):65-70

145. Lamb D. Could simulated emergency procedures practised in a static environment improve the clinical performance of a Critical Care Air Support Team (CCAST)?: A literature review. Intensive and Critical Care Nursing 2007; 23(1):33-42

146. Michelson J. Simulation in orthopaedic education: an overview of theory and practice. The Journal of Bone and Joint Surgery 2006; 88-A(6):1405-1411

147. Rosenbaum ME, Ferguson KJ, Lobas JG. Teaching medical students and residents skills for delivering bad news: a review of strategies. Academic Medicine 2004; 79(2):107-117

148. Seymour NE. VR to OR: A review of the evidence that virtual simulation improves operating room performance. World Journal of Surgery 2008; 32(2):182-188

149. Undre S, Darzi A. Laparoscopy simulators. Journal of Endourology 2007; 21(3):274-279

150. Villegas L, Schneider BE, Callery MP, Jones DB. Laparoscopic skills training. Surgical Endoscopy 2003; 17(12):1879-1888