Evaluation of simulation-based education in the management of medical emergencies.

Jennifer Mary Weller

This thesis is submitted in fulfillment of the requirements of the degree of Doctor of Medicine, The University of Auckland, 2005.

Abstract

Evaluation of simulation-based education in the management of medical emergencies.

Introduction

The traditional approach to medical education is changing and simulation is increasingly being incorporated into the curriculum, particularly in the context of emergency care.

Simulation takes many forms, but this thesis refers only to whole patient simulators, where a computerised mannikin placed in a clinically appropriate environment is used to recreate a realistic clinical encounter. The focus of this work is the management of medical emergencies.

Aim

The overall aim of this body of work is to evaluate the effectiveness of simulation-based education across a range of different learners, and to investigate the properties of simulation-based assessment.

In particular:

- To review the literature on the effectiveness of simulation, and its use in assessment.
- To evaluate the current status and effectiveness of CME interventions and the relative usefulness of simulation courses in this context.
- To determine if simulation-based courses in crisis management can lead to changes in physician behaviour.

- To evaluate student perceptions of learning in a simulator environment.
- To assess students' ability to manage simulated emergencies, and their opinion of simulation-based assessment.
- To define the psychometric properties and feasibility of simulation-based assessment in anaesthesia, the accuracy of self-assessment and the impact on learning.

Methods and Results

A number of different methods were used, which will be described in detail in the subsequent chapters. Overall, the results provide evidence for the effectiveness of simulation across a range of applications. Simulation-based assessment is acceptable, likely to have a positive impact on learning, and evidence support aspects of validity. Reliable scores can be generated but large numbers of cases are required.

Conclusions

There is sufficient evidence to recommend incorporating simulation-based courses into the acute care curriculum of medical undergraduates. Simulation is effective in CME in the context of anaesthesia crisis management and this is likely to apply to other acute care specialties. Like other clinically based assessments, extended testing time is required to generate reliable scores, limiting the feasibility of large scale, simulation-based exit examinations.

Preface and Acknowledgements

This thesis is a synthesis of the following series of publications around the theme of simulation in medical education:

Weller, J. and A. Woodward, Continuing Medical Education: What for? How? And how much is it worth. New Zealand Medical Journal, 2004. 117(1193): p. U876.

Weller, J. and M. Harrison, Continuing education and New Zealand anaesthetists: an analysis of current practice and future needs. Anaesthesia and Intensive Care, 2004. 32(1): p. 59-63.

Weller, J., L. Wilson, and B. Robinson, Survey of change in practice following simulation-based training in crisis management. Anaesthesia, 2003. 58(5): p. 471-473.

Weller, J., A. Dowell, M. Kljakovic, and B. Robinson, Simulation training for medical emergencies in general practice. Medical Education, 2005. in press.

Weller, J., Simulation in undergraduate medical education: bridging the gap between theory and practice. Medical Education, 2004. 38: p. 32-38.

Weller, J., B. Robinson, P. Larsen, and C. Caldwell, Simulation-based training to improve acute care skills in medical undergraduates. New Zealand Medical Journal, 2004. 117(1204).

Weller, J., B. Robinson, B. Jolly, L. Watterson, M. Joseph, S. Bajenov, A. Houghton, and P. Larsen, Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores. Anaesthesia, 2005. 60(3): p. 245-50.

I would like to acknowledge the contribution made by my co-authors, and in particular by Dr Brian Robinson, the Director of the National Patient Simulation Centre, without whose support a number of these studies would not have been possible.

I would also like to acknowledge my colleagues who assisted me during the writing of this thesis. Professor Iain Martin who provided invaluable advice on each chapter as it was written, and Ms Dulcie Brake for her help in proof-reading and formatting the final document.

Finally I would like to acknowledge the suggestions and encouragement provided by my husband, Professor Alistair Woodward.

List of Tables

List of Figures

Chapter 1 Evaluation of simulation-based education in the management of medical emergencies

1.1 The changing landscape of medical education

Traditional medical education has relied on formal instruction in basic sciences and an apprenticeship in clinical practice. A number of factors are changing the face of both medical practice, and medical education and new approaches may be required to meet the future needs of doctors and the society in which they work.

In the past, medical schools have described their curricula in terms of lists of course contents. Learning objectives then became popular, focussing teachers on what it was students should learn. The move now, endorsed by the World Federation of Medical Education, is to describe the curriculum in terms of learning outcomes, or what students can do when they graduate, and to ensure assessments are designed to measure these outcomes [1, 2].

In America and the United Kingdom, national bodies are taking a lead in this transition. The United States Accreditation Council for Graduate Medical Education (ACGME) aims to improve resident education and assessment through an emphasis on learning outcomes, based on competencies in six areas: patient care, medical knowledge, practice-skills, professionalism and systems-based practice. The "ACGME Toolbox of Assessment Methods" provides a framework for the development of discipline specific assessment of competencies [3].

The United Kingdom General Medical Council, in Tomorrow's Doctors, also frames medical education in terms of learning outcomes, and medical schools must demonstrate how their assessments ensure these curricula outcomes are achieved [1].

There are additional drivers in medical education suggesting change is needed in current medical school programmes. Increasingly there is a move to base educational practice on the best evidence available. Kaufman [4] proposed a number of principles to guide teaching practice. These include actively engaging the learner, solving real-life problems, providing opportunities for practice, giving feedback, and facilitating reflection on practice. Learning occurs where students engage actively with the task provided, and is consolidated by in-depth examination of the new experience. Students are more likely to learn material that is perceived as relevant, and more likely to be able to retrieve information when it is learnt in the context in which it must be applied. Many medical schools have responded by integrating case-based or problem-based learning through their curricula, and establishing departments of medical education to promote more effective teaching strategies.

Much is written in clinical education journals and workforce planning documents on the need for interdisciplinary education, improved communication skills in health professionals, and the ability to work in teams [5, 6]. Opportunities for meaningful, clinically based interdisciplinary learning are often limited in traditional medical education, as are opportunities to develop teamwork and communication skills.

With the rapid expansion of medical knowledge, and increasing attention to maintaining professional competence, skills in self-directed, life-long learning have become increasingly important attributes for new graduates. Educational methods that promote reflection, facilitate feedback on practice, and assist students to identify their own deficiencies and provide an opportunity for deliberate practice are likely to encourage self-directed learners, who can recognise their own learning needs and maintain competency throughout their lifetime as clinicians [7-9].

The focus on quality and safety in health care is also influencing trends in medical education. Reports from four countries identified an alarming rate of medical error and iatrogenic harm in hospital patients [10-13]. Deficiencies in management of the critically ill patient contribute to this iatrogenic harm. In addition, Human Factors research has offered insight into the underlying causes of medical error and why doctors make mistakes [14-21].

In the past, there has been a heavy reliance on the hospital apprenticeship to ensure medical students developed the required clinical competencies. In the modern hospital system, there are now fewer opportunities for medical students and junior doctors to gain this clinical experience as patient hospital stays are shorter, and fewer patients are admitted to hospital, resulting in fewer patients available for students and resident doctors. This is compounded by a reduction in working hours for resident doctors in many countries, including New Zealand.

The ethical landscape has also changed and the community does not expect to be used as practice models. There are well documented learning curves for procedural skills, and where alternative models are available, it would seem unethical to learn for the first time on a real patient [22].
In the past, it was not uncommon to learn emergency procedural skills, for example, insertion of central venous lines or cricothyroidotomy, on the recently deceased. Except in the unusual situation where prior consent has been obtained, this is no longer acceptable practice [23].

The use of live animals for learning physiology and invasive techniques is also problematic and appears to be declining in the medical undergraduate curriculum [24]. Maintaining an animal laboratory, which includes the cost of increased security, is a major financial investment, and not all students find the use of animals for the purpose of learning acceptable [25].

As a result, the traditional medical programme may not be up to the task for producing medical graduates with all the required competencies, in particular in the domain of acute care.
Increasingly, computerised patient simulators are being used to overcome the difficulties of learning on real patients.

1.2 Simulation in medical education

Simulation as a method of teaching is not new, and simply refers to the use of a model or person to simulate a clinical procedure or patient condition for the purposes of training. Simulated patients, standardised patients or peers are widely used as patient substitutes. Screen based simulations of clinical cases are available with varying degrees of sophistication. Static models have long been used for teaching basic practical skills, and there is an increasingly complex array of endoscopy and surgical procedural simulators. A range of full or partial body simulators are available for specific purposes, for example the Laerdal range of resuscitation and airway trainers, and the cardiovascular trainer “Harvey”, which is capable of displaying realistic cardiovascular signs with accompanying visual displays.

The more sophisticated of the whole body patient simulators enable immersion in a realistic clinical situation. In an ideal world, a complete replica of the clinical environment would be available for educating new doctors, providing a totally realistic but safe learning environment for training and assessment. The airline industry approaches this ideal with aviation simulators, and pilots may be certified to fly an aircraft with no training outside of a simulator. As it is easier to replicate an aircraft cockpit than a human body, it is unlikely that this will be achieved in medicine in the near future. Increasing simulator fidelity comes with a cost that limits feasibility, and gains in terms of educational benefit may not justify the additional expense.

The first computerised whole body patient simulator was built in the 1960s [26], but technology at the time was inadequate for the task. The computer required to run the simulator took up an entire room which meant it was not a feasible proposition for widespread use. It took another 20 years for a commercially viable patient simulator to become available, and its use in medical education was first described by a group of anaesthetists in Stanford [27, 28]. Anaesthesia has led the field in simulation-based education, but increasingly other disciplines have become involved, and patient simulators are now used widely in a number of contexts at both undergraduate and postgraduate level.

The patient simulators that are currently available have a realistic upper airway, chest movement, a range of cardiac and breath sounds and palpable pulses. The mannikin can be ventilated, intubated, cannulated, catheterised, given fluids and medications, paced and defibrillated. Needle thoracocentesis and chest drain insertion are possible. Monitors display blood pressure, ECG, oxygen saturation and expired carbon dioxide. Clinical signs and monitored data change in response to the evolving clinical event and participant interventions.

The most sophisticated simulator currently on the market is the METI (Medical Education Technologies Incorporated) Human Patient Simulator. Its physiological and pharmacological computer modelling can respond realistically to changing states or interventions, such as hypoventilation, blood loss, wrong drug administration or myocardial infarction. The simulator lungs exchange oxygen, carbon dioxide and volatile anaesthetic agents, which can be measured with standard gas analysis monitors. When connected to a patient ventilator, the METI simulator generates realistic airway pressures, tidal volumes and spirometry traces. It can be connected to real medical monitors, and in addition to displaying vital signs, allows measurement of cardiac output, blood gases and lung function.

1.3 Where does simulation fit in the broader picture of health education?

A range of teaching methods is required to ensure new graduates achieve the desired range of learning outcomes, and to ensure practising clinicians maintain their competence. There are currently some deficiencies in the existing medical education curriculum, and patient simulators could be used to address a number of these deficiencies through new course initiatives, or by improving the quality of current teaching programmes. There are some instances where patient simulators could increase efficiency of training and shorten times for acquisition of new skills and competencies.

There may be scope to assess some of the desired competencies of new graduates using a patient simulator. Students can be observed in highly standardised clinical scenarios, managing events that may be difficult or impossible to observe in real life. Teamwork and communication can be observed and assessed under a range of clinical conditions. Patient simulators, used in combination with other measures, may also have a place in assessment of clinical competence both in certification and recertification of doctors, and potentially in diagnosis and remediation of poorly performing practitioners.

The use of patient simulators in clinical teaching is based on recognised educational principles. Incorporating patient simulation into a course can increase clinical relevance, engage students in problem solving and decision making, and provide opportunities for feedback and repeated practice. Learning occurs in a clinical context, facilitating subsequent recall in the clinical setting.

Clinical events played out on a patient simulator can act as a trigger for clinically relevant multidisciplinary learning, increasing understanding of the perspective of other health professions. This provides an opportunity to learn more effective teamwork and communication, and facilitates exploration of the factors which affect performance in the stressful conditions of medical emergencies. Working through a simulated clinical case can help students identify gaps in their learning, as they are forced to make decisions and manage problems. Identifying gaps in learning is an essential precursor to the self-directed learning required to address these deficiencies.

Patient simulation provides a safe learning environment for the novice, where events can be scheduled, repeated and observed, offering the potential for greater efficiency and rigour over traditional methods, where a student’s clinical experience can be variable and ad hoc. A degree of clinical competence can be achieved prior to patient exposure. Learning new procedures requires practice and feedback and simulators may decrease training time. Simulators provide a unique opportunity to practice for rare or life-threatening events for students, doctors in training and experienced specialists.

Most clinical learning is best undertaken in a real clinical setting, where there are no limits to fidelity. However, in each domain of clinical medicine and assessment there are likely to be some areas that would be best taught using a patient simulator, and others that would benefit from the addition of patient simulators to their traditional courses.

However, in spite of its theoretical advantages, simulation is not widely used in New Zealand medical schools and is not a regular component of post-graduate medical training or continuing professional development. The aim of this thesis is to map the use of patient simulators in medical education, to determine the effectiveness of simulation-based medical education, to explore the properties of assessments using patient simulators and to define the place of patient simulation in undergraduate and post graduate medical education.

Chapter 2 Patient simulators in medical education: a review of the literature

2.1 Introduction

Abrahamson's report of the first whole body, computerised anaesthesia simulator appeared in 1969 [29], but not until the 1990s did a commercially viable model become generally available. Since then the body of literature on simulation has steadily increased, accompanied by a rapid expansion in the number of centres of medical simulation worldwide [30]. In 1998 there were an estimated 70 simulation centres around the world, compared with 20 in 1996 [31] and this had risen to 158 by 2001 [32]. In 2005, the Society in Europe for Simulation Applied to Medicine (SESAM) estimated there were 260 simulation centres in the United Kingdom alone. In New Zealand there are at least seven simulation and skills centres possessing a whole body computerised patient simulator.

For the purposes of this literature review, relevant databases were systematically searched for all articles on computerised full body patient simulation in medical education published between January 1990 and December 2004. A number of these publications explored the educational rationale for using simulators in medical education, described different types of simulators, offered guidance on how to set up a simulation centre, described how to incorporate simulation into the undergraduate curriculum or were limited to opinion [33-42]. These were not considered further. Articles with original material were divided into those describing different uses of simulators in medical education, those evaluating the effectiveness of simulation-based courses, and those exploring the use of simulation in assessment. This chapter will map the use of simulation in medical education, examine the evidence for its effectiveness, and outline the current use of simulators in summative assessment.

2.2 To what extent is simulation being used in medical education?

Simulators are used at all levels of medical education, from medical school to specialist continuing education programmes. The contexts in which they are used range from basic sciences tutorials to courses in crisis management.

2.2.1 Undergraduates

Simulation has been used in the medical undergraduate curriculum for teaching physiology and pharmacology, basic anaesthesia skills, and management of the critically ill patient.

2.2.2 Physiology and pharmacology

The most sophisticated simulators incorporate programmable computer modelling of physiological responses to a comprehensive range of clinical situations or interventions. Signals linked to standard monitoring equipment can generate graphical displays of vital signs, ECG, arterial, venous and cardiac pressures, cardiac output, airway pressure and flow-volume loops, end tidal and blood gas analysis, oxygen saturation and temperature. With this information, physiology can be brought to life through clinical scenarios played out in real time, demonstrating, for example, cardiovascular response to shock [40, 43-46], changes in the Starling curve in response to haemorrhage and volume loading [47], or effects of changes in ventilation/perfusion ratios in different clinical situations [43]. Patient factors can be incorporated into the design, for example to demonstrate the differing response of the failing heart and the healthy heart to volume loading. The use of simulation in this context has, in some centres, replaced live animal physiology laboratories [40, 43, 45]. The physiological response to any drug can be programmed into the simulator based on empirical data, and this capability has been utilised to teach pharmacology, for example to demonstrate inspired and expired concentrations of volatile anaesthetic agents, and the real time effect of these agents on the cardiovascular system [48].

2.2.3 Transition to clinical practice

Simulators have been used to help medical students acquire basic examination skills, interpret patient data, learn technical skills in vascular access, airway management and cardiac resuscitation, and learn how to manage the critically ill patient [40, 49-52]. In some centres, simulation has been incorporated into the curricula of surgical, anaesthesia and emergency medicine clerkships [51, 53-55].

2.2.4 Resident doctors

A number of post-graduate training programmes have looked at including a simulation component. Surgical interns have learnt aspects of critical care [56] and trauma assessment skills on simulators [57]. Simulators have been used in Advanced Trauma Life Support courses [58] and in airway management courses for hospital residents [59].

In anaesthesia, novice anaesthetists have been taught the basics of anaesthesia practice, including the use of complex equipment, using patient simulators [50, 60] and a number of studies describe courses in crisis resource management for anaesthesia trainees [61, 62]. The concept of crisis resource management originated in the airline industry, where it is referred to as Crew Resource Management [63], and in the early 1990s this concept was transferred to medicine in the form of Anaesthesia Crisis Resource Management (ACRM) courses [64]. The impact of factors such as leadership style, clarity of communication, situational awareness and resource allocation on patient management can be explored through realistic simulations of critical events. In some countries, ACRM-type courses are now a requirement for accreditation [61, 65].

Simulators are being used in training for emergency medicine [66-69]. For example, in the University of New Mexico, Albuquerque, a simulation-based course has been incorporated into the three year emergency medicine training programme, beginning with a graded introduction to core competencies, and moving to increasingly complex clinical scenarios [68]. A course in Emergency

Medicine Crisis Resource Management is described for emergency medicine trainees, who received it with enthusiasm [67].

The Noelle birthing simulator was developed for training obstetric residents. It can simulate both the technical aspects of operative delivery and an entire obstetric emergency scenario [70]. A simple neonatal resuscitation mannikin has been used in conjunction with a high fidelity maternal simulator to train paediatric residents to manage neonatal emergencies [71].

2.2.5 Medical Specialists

Most reports of patient simulators in continuing education of medical specialists address the management of medical emergencies [27, 50, 64, 72-77]. These courses are well received by participants [64, 74-76] and have been formally recognised in maintenance of professional standards programmes [72]**.** The majority of these reports are in the field of anaesthesia.

Simulation-based crisis management courses have also been described for rural general practitioners [52], for radiologists [78], and for military physicians [79], to maintain competence in management of uncommon, life-threatening medical emergencies.

In addition to training doctors from a single specialist group, there is increasing use of simulation to train multidisciplinary health care teams, again mainly in the context of crisis management [50, 72]. As well as broadly based courses on the generic principles of crisis management, there are descriptions of courses targeting specific health care teams, in particular in the fields of intensive care, paediatrics, emergency and retrieval medicine, hyperbaric medicine and the armed forces [72]. Simulators have been used in training teams of obstetricians and anaesthetists managing obstetric emergencies [50], operating room teams, and adult and paediatric trauma care teams [50, 80, 81].

A number of reports address responses to the effects of weapons of mass destruction, and chemical and biological warfare. A simulator has been developed that can portray some of the effects of exposure to chemical toxins, for example, excessive lacrimation and pulmonary oedema. Evaluations for this form of training have been positive and participants report an increased understanding of the limitations of providing medical care in adverse conditions, for example, inserting an intravenous cannula whilst wearing a full suit of protective clothing [82, 83].

Outside of acute care medicine, a high fidelity simulator interfaced with a computerised pharmacokinetic display of drug concentrations has been used to introduce new drugs, for example remifentanil, into anaesthetic practice [84]. The introduction of new equipment into anaesthesia practice using patient simulators has also been described [85], for example, learning how to use a new anaesthetic machine in a simulation centre prior to using it in clinical practice.

Several of the studies described above incorporated distance learning into the programme [48]. Medical students have observed the cardiovascular effects of volatile anaesthetic agents by relaying real time effects from a simulated operating room environment to a distant classroom [86]. Military teams in remote locations have used simulators controlled by an instructor in a central location and connected by video-link, to practice for medical emergencies and ensure ongoing competence while on extended periods of duty [79]. Interactive videoconferencing of life-threatening clinical problems relayed to an audience at a distance, using a combination of an actor and a computerised patient simulator, has also been described and was rated more highly by the audience than the traditional case-conference [87].

2.3 What evidence is there on the effectiveness of simulation-based education?

A number of methods can be used to evaluate the effectiveness of educational interventions. Ratings of perceived value of the intervention, self-reported increase in confidence, perceived improved performance, or subsequent reports of change in practice can give a measure of course effectiveness from the perspective of participants, but rely on accuracy of participant reports and self-assessment. Pre and post intervention testing is a frequently used technique, where care must be taken to ensure the test is appropriate to the desired learning outcomes, and potential confounders are taken into account (e.g. concurrent learning, and the potential for improved performance being due to increased familiarity with the simulator). Measuring actual changes in practice would ensure that the intervention not only improved test scores, but improved subsequent clinical performance. The optimal outcome measure in medicine is improved patient outcomes, but due to the multiple factors contributing to patient outcomes, it would be difficult to assign a change in morbidity or mortality to a particular educational intervention. Incorporating an appropriate control group is often problematic for all approaches to evaluation research in education.

Student ratings of the simulation-based courses described above have been enthusiastic [43, 44, 47-49, 51, 52, 54, 55, 86, 88, 89], and simulation workshops have been rated more highly than other teaching methods [47, 55]. A subset of the preceding publications include some measure of course effectiveness, over and above participant satisfaction and preference (Table 1).

Table 1: Effectiveness of simulation courses

Study	Outcome measure	Results	Comment
Euliano 2001: 68 medical students	Pre and post test of written knowledge. Confidence	Improved test scores. Improved confidence.	Cardiovascular physiology. No control
Tan 2002: 210 medical students	Pre and post test of written knowledge	Improved test scores.	Cardiovascular physiology. No control
Morgan 2002, 144 medical students	Improved performance on simulator	No difference between two interventions.	Basic anaesthesia skills. Randomised, Controlled trial. Controlled for simulator familiarity.
Nackman 2003: 54 medical students	Performance in “Shock stations” in surgical OSCE	Improved OSCE scores in simulation group	Management of shock. Historical control.
Cleave-Hogg 2002: Gordon 2001: medical students	Written comments from students	Identify knowledge gaps, increase confidence, theory to practice, taking responsibility.	Post course questionnaire.

2.3.1 Undergraduates

All reported studies using simulation in the context of undergraduate basic sciences have shown high levels of student satisfaction, and where measured, self-reported confidence has increased. Pre and post-intervention tests of students' written knowledge were used in studies by Euliano and Tan [44, 47]. In Euliano's study 78% of 87 students completed both pre and post-intervention tests, and in Tan's study all 210 students completed both tests. Both authors reported significant increase in scores for most questions following the simulator lessons, but there was no control group in either study. Euliano also enquired about students' confidence in their answers to questions on cardiovascular physiology, and found that in addition to improved test scores, students' confidence in their answers improved significantly.

Two studies have used improved performance on the simulator as a measure of effectiveness, and both included a control. Morgan et al [55] conducted a randomised, controlled study of 144 medical students and assessed basic anaesthesia skills following either an interactive workshop with a patient simulator or a faculty-facilitated instructional video on the same material. Performance was assessed on a patient simulator immediately before and after the workshop. The authors controlled for the effect of varying familiarity with the simulator by exposing the control group to a simulation-based workshop on an unrelated topic. In addition to the simulator assessment, all students sat a written assessment at some point between two and thirty days after the intervention. Both groups improved from pre to post-test on the simulator and there was no significant difference between the groups in either the simulator or written assessments. The authors concluded that they could not demonstrate an advantage of the simulator workshop over the alternative teaching session, but commented that future studies should include testing retention of knowledge over a longer period of time, and should compare modalities that were less similar.

Nackman [53] tested the hypothesis that exposure to a one-hour teaching session incorporating patient simulation would improve medical students' performance in the end of run "shock stations" of the surgical objective structured clinical exam (OSCE), when compared to a historical cohort who received a traditionally structured, case-based session of similar duration. Performance was assessed using the RIME (Reporter, Interpreter, Manager, Educator) model [90], grading the ability of students to collect information, generate a differential diagnosis and prescribe appropriate management. Fifty-four students from two different hospitals in the same training programme were enrolled in the study, and were compared with students in the preceding year prior to the availability of the simulator. The authors controlled for hospital site, time of year, and previous examination scores and found that site of study and exposure to simulation were the only independent predictors of OSCE performance. OSCE scores were significantly higher in the simulator group. They concluded that the simulation session was more effective than the traditional lecture and improved the ability of students to assess shocked patients. Limitations of this study include lack of randomisation and use of a historical control. In addition, there is often instructor enthusiasm with the introduction of a new teaching method, and this may have resulted in increased motivation of students. Improved performance due to the enthusiasm of the experimenter or teacher rather than the specific intervention is known as the Hawthorne effect[1], and is a common confounder in educational research.

In addition to this limited quantitative evidence, several studies have sought students' written responses to open questions evaluating simulation-based workshops [47, 51, 54]. Cleave-Hogg and Morgan [54] took groups of ten students through a simulator workshop where each was exposed individually to a short case based on the anaesthesia curriculum (total of 145 students). They were watched by their peers and provided with feedback. Gordon et al [51] designed a workshop where 27 students were individually mentored by an educationalist through a scenario depicting a shocked patient. A number of common themes emerged from the student evaluations of these two courses. Students valued the active engagement with the case, the opportunity to take responsibility for

[1] Improvement in the performance of workers at the Hawthorne factor of the Western Electric Company in the USA was assumed to be an effect of specific changes made in working conditions, (eg in the level of illumination), whereas it was eventually established that the workers were responding to the interest shown towards them by the investigators. (Cambridge Encyclopaedia 1990)

patient management decisions, and the chance for hands on practice without the risk of patient harm. They felt the simulations helped identify gaps in their knowledge, highlighted what they needed to know, and increased their confidence. Students also appreciated being able to apply their theoretical knowledge to a clinical situation.

2.3.2 Resident doctors

A number of studies have evaluated the effectiveness of simulation in the context of management of critically ill patients. They demonstrate some of the difficulties of evaluating educational interventions at this level of complexity.

Using improved simulator performance as the outcome measure, Marshall et al [58] evaluated the effectiveness of incorporating trauma scenarios on a patient simulator into an Advanced Trauma Life Support course (ATLS). Using a pre and post-test design, they scored performance in two initial trauma simulations and compared this with performance in two simulations at the end of the course. Scores increased from pre to post tests, and participant self-confidence was significantly improved. However, no control group was included in this study and numbers were small (12 residents). It is difficult to know how much reliance can be placed on the test scores as there is limited information on the assessment criteria, performance was measured in only two scenarios before and two after the course, and only two raters scored the performance. There is no data given on the reliability of the scores, although construct validity was supported by demonstrating superior performance in a group of more senior residents.

Lee et al [57] conducted a prospective, randomised trial comparing the effectiveness of actors made up with moulage to resemble trauma victims, and trauma scenarios using the patient simulator, in training residents in the early management of severe trauma. Following an ATLS course, 60 interns at two sites were randomised either to a practice trauma scenario with an actor, or a practice session with a simulator, and then randomised again for testing either with actor or simulator. Nurses assisted the residents during the scenarios. Residents were scored by two judges, one during the scenario, and one at a later time by video, using a crossover design for the judges. Lee found mean trauma assessment test scores were higher for all simulator-trained residents when compared with actor-trained residents, and simulator training independently showed a small but significant improvement in the total score and event score. He concluded that use of a patient simulator was feasible and compared favourably with training in a moulage setting with actors. Reliance on the performance scores is in question due to the large reported error associated with the judge, and the additional error due to variable prompting by the nurse. Furthermore, the assessment consisted of a single case, and performance in one case may not be a reliable predictor of performance in other cases (context specificity). However, it could be argued in this study that the assessment criteria were generic, consisting of a checklist describing the steps in the primary and secondary survey of the trauma victim, and as such the score could be independent of the specific clinical context.

Mayo et al [59] tested airway management skills of new interns, and attempted to correlate simulator training with clinical performance. Following initial testing on the simulator, interns were randomised to further simulator training or no training, then tested again on the simulator four weeks later. Those who had received training gained significantly higher scores. The “no-training” group then participated in the simulator training sessions. The authors then scored airway management of these interns during emergency calls over the subsequent ten months, and found very high standards of performance, which they attributed to the simulator training (41 of the 50 residents were observed at least once during an emergency call). This study demonstrates the difficulty of using a control group in clinical education. The authors compared simulator training with no training, and when looking for transfer of simulation training to the clinical environment, used a historical control, i.e. anecdotal evidence that interns’ airway skills in emergency calls had been deficient prior to the introduction of simulator training. However, this study does attempt to incorporate a very relevant outcome measure, i.e. performance in real life emergency calls, rather than the proxy measures of improved scores in written tests or on the simulator.

2.3.3 Specialists and specialist trainees

Sica et al [78] studied the effect of a half day simulation workshop on crisis management skills of 24 radiology residents. They demonstrated significantly improved performance in the second simulation at the end of the workshop. This was a pre-post test design, with limited numbers and no control intervention. Improvement could potentially have been due increased familiarity with the simulator and again, the reliability of the assessment is difficult to judge.

Chopra et al [73] investigated the long term effect of simulator training in the management of malignant hyperpyrexia (MH) on subsequent scores in a simulator assessment, controlling for familiarity by including a group who trained on the simulator with a different clinical scenario. They recruited 28 anaesthetists and trainees to the study. Performance scores in initial baseline simulation of anaphylaxis were used in an attempt to adjust for variation in subsequent scores due to differing ability of the participants. Participants then trained either on anaphylaxis or MH, and were then tested on MH four months later. Scoring was against a weighted checklist of specific actions, as well as times to key actions and deviations from recommended treatment protocols. The group trained in MH performed significantly better in the MH test scenario than the group trained in anaphylaxis, and showed a significant improvement in all scores from their first MH scenario. However, numbers were small, there was a single rater, and controlling for varying ability in MH using scores in an anaphylaxis scenario may not be valid in light of the known variation in an individual's performance between cases (context specificity). Nevertheless, this study suggests a sustained improvement in performance in managing a specific clinical event, which was not due to increased familiarity with the testing instrument.

Nyssen et al [62] used much of Chopra's methodology in a complex study comparing screen based computer simulations and patient simulators for teaching management of anaphylaxis. To allow for increased familiarity as a cause for improved performance, they included a control group who trained on a different scenario (malignant hyperpyrexia), and attempted to control for varying ability of the participants by choosing trainees with similar results in a previous anaesthesia examination, and excluding those with clinical experience of anaphylaxis. Forty anaesthesia trainees were randomised to two groups: Group A = anaphylaxis training and anaphylaxis test, Group B = malignant hyperpyrexia training and anaphylaxis test. These groups were further divided into two groups: screen based and patient simulator. All participants were tested on anaphylaxis four weeks later, using the modality on which they had trained. Two raters scored all anaphylaxis scenarios against the same highly specified checklist used in Chopra's study and the interclass correlation coefficient of two raters was found to be high (ICC= 0.99). Mean scores improved in Group A (trained and tested on anaphylaxis) from initial testing to repeat testing. Training modality (computer or patient simulator) and length of anaesthesia experience had no significant effect on the test scores. The latter finding calls into question the validity of the scoring, as more experienced residents would be expected to perform better, and the authors suspected prior knowledge of scenario content in the novice group. Improved scores did not appear to be due to increased familiarity with the simulator. The authors concluded that screen based simulations may be as effective as patient simulators in teaching technical skills and are more affordable. This study did attempt to include a control intervention, and attempted to control for increasing familiarity and varying participant ability. However, there was considerable variation in an individual's test scores, and combined with the small numbers, the reproducibility of these results may be questioned. Furthermore, it is unclear to what extent the different modalities would transfer to real life performance. The authors suggested that the more complex aspects of crisis management may be better addressed using patient simulators than computer screen simulations.

Byrne et al [91] evaluated the added value of video replay in simulator training sessions on anaesthesia emergencies, using time to solve the problem and chart recording error as the outcome measures, the latter being used as a measure of spare workload capacity. Participants were divided into two groups: one group was shown the video of the simulation during the debriefing, and one group was debriefed without reference to the video. The groups managed a series of different emergency scenarios over the course of the workshop, each followed by debriefing. Byrne found no significant difference in scores for the two groups, or between the first and last scenarios in either group. Because of the non-specific performance measures, the fact that all scenarios were different, and errors due to the reported variability in chart recording, it is difficult to draw any conclusions from this study regarding the value of simulation or the added value of video replay in the debrief. However, this study does attempt to measure the effects of comparable educational interventions, and illustrates the use of novel outcome measures.

A further difficulty in evaluating effectiveness of training was demonstrated by Olympio [60]. In a retrospective review of videotapes of 21 anaesthesia residents in simulator scenarios, he found no benefit from teaching a particular approach to management of oesophageal intubation. It appears that the lessons taught in the simulation centre were not supported by teachers in the clinical environment, which may explain why residents did not adopt the approach. This demonstrates the need to ensure simulation instructors and clinical teachers are both teaching to the same curriculum, and lessons learnt in the simulator are reinforced in the clinical teaching environment.

Training with whole body simulators has generally been evaluated in the context of emergency care. However, Murray et al [84] evaluated the effectiveness of learning how to use new anaesthetic agents using a METI simulator integrated with a computer screen displaying remifentanil concentrations. Pre and post determinations of level of comfort using remifentanil, and perception of changes in practice regarding remifentanil after participation were sought from 836 participants, with a 68.6% response rate. There was a significant increase in comfort level, and the training was perceived by respondents as a useful addition to available teaching methods, having the advantage of real time visual display of drug concentrations linked to clinical events portrayed on the simulator.

Dalley et al [85], studied the impact of simulation training on the introduction of new equipment into anaesthesia practice, using as an outcome measure, the number of errors committed by anaesthetists when using the new machine in a simulated case involving a clinical problem. Participants, who were all familiar with the simulator prior to the study, were randomly assigned to the traditional training or simulator group. The traditional training process of familiarisation comprised a didactic presentation by the company representative, demonstration of the features of the machine and an opportunity to ask questions and explore the features of the machine. The simulator group were also told about the machine by the company representative but had an opportunity to use the machine in a simulated routine case. Following training, all participated in a standardised test scenario on the simulator, which was videotaped and analysed for clinically important errors in the use of the machine. Significantly fewer errors were made in the simulator group. This study used an appropriate control group (although training time was different between groups), and used an appropriate outcome measure, i.e. errors committed using a new anaesthetic machine under adverse clinical circumstances. It would be difficult to test ability to use new equipment under adverse conditions in the operating theatre, and simulation may be a valuable modality for ensuring anaesthetists are competent in the use of new equipment in different clinical circumstances prior to using it in real life. There is also the potential to road-test new equipment under a range of clinical conditions in the final stages of its development.

2.4 Discussion

The perfect evaluation study of simulation is difficult to conceive. Most studies rely on proxy measures, and suffer from a number of constraints and potential sources of error.

A number of different outcome measures were used in the studies described above, and were generally not ideal. Increased confidence does not equate with increased competence, self-reported changes can be inaccurate, and self-assessment of one’s own performance generally correlates poorly with external assessments [92]. Improved performance on the simulator may not predict improved performance in the clinical setting and even observed clinical performance does not necessarily predict habitual practice. Only one of the studies described above attempted to correlate performance in the simulator with clinical performance [59], but due to the absence of a control group, no firm conclusions could be drawn.

There are anecdotal reports from course participants of improved patient outcomes as a result of prior training on the simulator, but this has yet to be demonstrated in a formal study, and even in the general medical education literature, there are few studies linking improved patient outcomes with educational interventions [93].

The studies also illustrate the difficulties and constraints around evaluation of educational interventions, for example, availability of students due to timetabling issues, curriculum constraints, service requirements and expense. The ability to test retention of learning, a potential advantage of simulation, was limited by availability of undergraduates outside of brief clinical attachments [55] or a single attendance at a course. There is also an ethical responsibility to students to ensure no group is disadvantaged by exclusion from an educational offering which is available to their classmates. If random allocation of students is required, attention must be given to the control group, for example by providing an alternative experience or by offering the intervention at the end of the study period.

Choice of an appropriate control is also problematic. A number of studies have compared the simulator intervention with the status quo, which was generally ad hoc clinical experience with no specific or formal training in the area under study. These studies demonstrated that the training was effective, but could not demonstrate that it was more effective than an alternative intervention. In the single study where simulation was compared to another innovative and interactive intervention [55], no advantage was shown for the simulation intervention. A simulation workshop has also been compared to a traditional lecture [53]. If equal time is given to both interventions, this may be an appropriate control group, as it reflects widespread teaching practices. Choice of a control group will ultimately depend on what the comparison is attempting to demonstrate. This may be the effect of the addition of simulation to an apprenticeship style training programme, comparative effectiveness of simulation-based interventions and traditional lectures and tutorials, comparison of simulation with an alternative teaching innovation, or comparison of alternative methods of delivery of a simulation-based educational intervention. The last approach would have the benefit of identifying ways of improving simulation-based teaching rather than further attempts to justify the use of simulation in the medical curriculum.

These studies illustrate the numerous sources of error that inevitably reduce the reliability of the reported results. Where interventions take place over a period of time, there is always the potential for concurrent learning outside the intervention, and the knowledge that some students are receiving instruction on a topic may prompt others to learn independently. The initial enthusiasm of the teacher for a new method may be the cause of improved learning, rather than the method itself. Sampling of participants was randomised in some studies, but in others a convenience sample was used, with the potential for bias. For surveys with less than 100% response rate, responders may differ from non-responders in their positive views of the simulation experience, and the ability to generalise results across courses, disciplines, institutions or national boundaries may be limited.

Quantitative evaluation of an educational intervention frequently involves assessment of performance. Where the outcome measure was improved performance, the validity and reliability of the assessment process was often difficult to gauge. This will be explored further in the next section.

2.5 Using simulation to assess performance

A number of studies described in the preceding chapter assessed performance in the simulator as a measure of course effectiveness. This chapter will review the studies where the primary goal was to evaluate the use of simulators in assessment (Table 2). To date these studies have focused on acute care skills and management of medical emergencies.

Table 2: Summary of assessment studies

IRR =Inter-rater reliability, IC =Internal consistency, ITA=In Training Assessment

Study	Measures of Reliability	Measures of validity	Scoring system
Murray 2002, 64 medical students	IRR IC	Construct; content; Correlation-checklist/holistic	Checklist & global score
Boulet 2003, 40 medical students	IRR, Inter-case	Construct, content	Checklist
Morgan 2000, 24 medical students	IRR	Content; correlation –simulator / written / clinical	Checklist
Morgan 2001, 135 medical students	IRR	Correlation-checklist /global simulator / written / clinical.	Checklist & global scores
Morgan 2004, 135 medical students	IC, Inter-case		Checklist
Gordon 2003: 23 mixed group		Construct; correlation - OSCE /Simulator	
Tsai 2003: 20 paediatricians	IRR, IC	Construct; content; Face	Checklist
Holcomb 2002: 15 Trauma teams	Inter case	Construct; content	Checklist
Forrest 2002: 13 new anaesthetists		Construct; content	Checklist
Murray 2004: 28 anaesthetists	Inter-case	Construct; correlation- global / checklist	Checklist & global score
Schwid 2002: 99 anaesthetists	IRR, IC	Construct; correlation- simulator / ITA /oral	Checklist
Devitt 1997: anaesthetists	IRR		Checklist
Devitt 1998: 25 anaesthetists	Internal consistency	Construct	Checklist
Devitt 2001: 146 anaesthetists	Internal consistency	Construct; face validity	Checklist
Byrne 1997: 20 anaesthetists		Construct	Time to solve problem
Gaba 1998: 14 anaesthesia teams	IRR		Global score technical/behaviour
Weller 2003: 28 anaesthetists	IRR	Content	Global score medical/ behaviour

2.5.1 Undergraduates/new graduates

One of the goals of curriculum reform in the United States is to ensure medical graduates acquire good patient management skills, and an important subset of these skills is management of the critically ill patient [3]. The potential to simulate medical emergencies led to the consideration of simulation-based assessments in licensing examinations, and to ensure these assessments were reliable and valid, research was needed to determine their psychometric properties.

Two studies by the same group undertook this task. In 2002, Murray et al [94] conducted a pilot study, and the following year Boulet et al [95] conducted a definitive study to investigate the psychometric properties of a simulation-based assessment in the domain of acute care skills.

In the pilot study, 64 medical students or residents participated in a five minute trauma scenario which had been developed by nine course conveners from the curriculum objectives. A scoring checklist was devised by three medical experts, which consisted of important and essential knowledge and skills expected of a graduating medical student. Each of the 18 items was scored as ‘yes’ or ‘no’, and weighted by importance. Three doctors and one nurse scored each of the study participants, using video review which included a display of patient vital signs. Two final scores from each rater were generated: the sum of credited items (0-18) and the weighted score (0-29). The mean candidate scores, both unweighted and weighted, were also calculated by averaging the scores from the four raters. Two additional experts independently awarded a holistic score for overall performance using a 10 cm visual analogue scale. The main criteria for this score were thought processes, action, and integration, each of which included explicit observable actions. The holistic score for each examinee was the mean of these two raters’ scores. The researchers found a high degree of correlation between the mean weighted and unweighted scores (Pearson correlation r=.87). They were able to assess each checklist item for difficulty and ability to discriminate between high and low performers. The inter-rater reliability for checklist and holistic scores were 0.92 and 0.81 respectively and there was moderate correlation between checklist and holistic scores

(0.66-0.71) for the four analytic raters against the mean holistic score), indicating that a moderate amount of the variance in holistic scores was explained by the number of checklist items attained. In general, more experienced examinees scored more highly. The authors concluded that multiple raters were not required to generate reliable checklist scores, and that these scores were a reasonable predictor of holistic scores generated by experts.

In a further study by this group using similar methodology, 40 medical students and new graduates were each scored in six of ten possible acute care scenarios developed by a group of specialists, and based on the emergency medicine curriculum [95]. There were two possible sets of scenarios, with two scenarios common to both sets. A weighted checklist was used, and scoring was conducted by two medical specialists and two nurse clinicians, each scoring independently. There was modest correlation between participant experience and level of performance, supporting construct validity of the assessment. Using generalisability theory, the investigators found that the variance in scores due to the rater was low, but there was a large variance due to the case, and concluded that acute care skills could be validly and reliably measured in this simulation-based assessment, but a large number of cases would be required. The authors commented that the high level of agreement between the raters was probably due to the highly specified items on the checklist, and prior agreement between raters on these items during a formal rater training session. This assessment was of junior doctors in clinical scenarios with a well defined management pathway. Where treatment pathways are less clear cut, and attempts are made to measure complex levels of performance in experienced practitioners, the use of highly specified checklists may not be as valid as holistic scores by expert judges.

In a subsequent review of the use of simulation in the United States Medical Licensing examination, Dillon et al [96] suggested patient simulators had a unique role to play. They could model rare events, and test clinical skills that were difficult or impossible to examine using real or standardised patients or actors. Real time responses could be measured, and scoring systems could be based on measurable "patient" outcomes, e.g. simulator death, which would be difficult or impossible to do consistently with real, standardised or actor patients. Importantly, patient simulators also provided an opportunity to assess the ability of a graduating doctor to work in a multidisciplinary team.

The second group of papers come from Morgan and Cleave-Hogg in Toronto, who focused on the use of patient simulators in undergraduate anaesthesia teaching and assessment. Their first study [97] sought staff and student opinion of the use of the simulator in assessing performance, and found both groups were, on the whole, supportive, but suggested that prior exposure to the simulator environment was required. This acceptance of simulation-based assessment was supported in a subsequent study [98] where students rated it highly, thought the assessment content was appropriate and considered it was a useful examination tool. In this study, 24 medical students each worked through six scenarios based on core objectives of the anaesthesia attachment, and were subsequently scored independently by five raters against a set of criteria. They found high inter-rater reliability with an intraclass correlation coefficient of 0.87, but simulator scores correlated poorly with both written examination results and clinical marks. The authors included reliability data for the written examination but not for the clinical marks, which consisted of faculty ratings of students following the anaesthesia attachment. Lack of correlation of these different assessments could mean either that one or more of the tests was unreliable or invalid, or that they measured different things. Correlation of 100% between tests would suggest that no additional information was gained and that one of the tests was redundant. Morgan commented that the students' clinical experience varied, and many students would not have been exposed in the operating room to the material in the simulations, which incorporated problems arising during anaesthesia, suggesting the assessments were measuring different attributes.

In a second, larger study of 135 final year medical students and five elective students [99], Morgan et al confirmed the reliability of checklist scoring, and the poor correlation between simulator scores, written test scores and clinical marks during the anaesthesia attachment. During their anaesthesia attachment, students participated in a videotaped simulator scenario containing five problems based on the published objectives of the undergraduate anaesthesia curriculum. The scenarios were tightly scripted and prompts were standardised. Two faculty members independently scored each videotaped scenario. Over the course of the year, ten faculty members were involved in the ratings, and were assigned to marking pairs, each pair marking three or four groups of students. The scoring was by a 25-point checklist where five points were awarded for each of five problems requiring specific actions. Inter-rater reliability was estimated for one and two raters and reported as 0.77 and 0.86 respectively and inter-item correlations in the simulations were variable. To achieve a reliability of 0.9, the authors estimated 2.68 raters would be required, using their test format. The authors did not consider alternative test formats as a way to improve reliability of the results. Based on analysis of comparable assessments using generalisability theory [100], it is likely that increasing the number of cases rather than the number of raters would do more to reduce the overall error in the scores [101]. Correlation of simulator scores with written and clinical marks was low, suggesting that the written and clinical assessments were not testing important objectives of the anaesthesia rotation, i.e. the ability to manage common problems arising during the course of anaesthesia.

The correlation between checklist and global scoring was also addressed in this study. In addition to the checklist scores, the faculty were asked to give a global score for performance of students [102]. Global scores correlated well with the checklist score (r=0.74), but had lower reliability. Single rater reliability for checklist was 0.77 (range 0.58-0.93) and for global scores 0.62 (range 0.40-0.77). The checklist scores were then divided into three domains: knowledge, technical skills and judgement, and the correlations with the global scores were 0.21, 0.51 and 0.53 respectively for these domains. The authors concluded that the checklist scores were more reliable than the global scores but both were acceptable, that there was good correlation between the two scoring methods, and this was higher in the domains of judgement and technical skills than knowledge. The latter finding could indicate the global scores were assessing higher levels of performance, but there are some limitations to the study design. Each pair of examiners marked one fifth of the tapes, which included six different scenarios. Test error could arise due to variable performance across the cases, the way a rater scored a particular case, or an interaction between the rater and a particular student. The high level of correlation between the checklist and global scores could have been because raters awarded both scores, and one score influenced their decision on the other.

Morgan and Cleave-Hogg subsequently explored the correlation between student scores in the previously described scenarios and their prior experience and confidence [103]. A single faculty member scored student performance in three scenarios using pre-defined checklists. Prior to the simulations, students rated their experience and confidence in 25 clinical items, eight of which were included in the scenario checklist. The authors found no correlation between student ratings for clinical experience and confidence, and the subsequent simulator test score. Level of confidence also showed no correlation with clinical grades or written examination marks.

These authors went on to design ten anaesthesia scenarios each containing a discrete problem, for example, hypoxaemia or dysrhythmia [104]. They created performance checklists for each scenario, and to increase the validity of these checklists, sought agreement between 15 faculty on the appropriateness and importance of each item. They then scored the performance of 135 medical students in the ten scenarios against this checklist. Thirty of these students participated in a second scenario, generating 165 scenarios for scoring. For these 30 students, no correlation was found between the scores in the two cases. Again there was no correlation between the simulator scores and the scores for written and clinical marks during the anaesthesia rotation. Internal consistency of the checklist items was measured using Cronbach's α, where a level of 0.6 or above was considered acceptable. Two of the scenarios were found to have acceptable internal consistency with all checklist items included, and an additional three scenarios were acceptable with one item deleted. This raises an interesting question. The checklist items had been chosen following a process of consultation with 15 faculty, and the final checklist items were those agreed on by 80% of these faculty members as being appropriate, and where it was considered removing them would have a negative impact on the validity of the examination. Where these items were subsequently shown to negatively impact on internal consistency, removing them improved reliability, but could reduce the validity of the assessment, and the extent to which it measured important attributes. This illustrates a general difficulty where reliability may be improved, but at the expense of the validity of the assessment. Poor correlation of an item with overall score may mean it is measuring something different and possibly important.

This group of studies is significant in that it represents progressive evaluation of simulator scenarios and performance checklists, provides information on inter-rater reliability and internal consistency, and identifies the poor correlation between simulator scores and other the methods of assessment.

The latter finding is supported by the results of the study by Rogers et al [105] who found that marked improvement in written test scores of 21 medical undergraduates following a critical care attachment was not matched by an equal improvement in OSCE and simulator test scores, where the material for each of the three assessments was based on the same clinical scenario. The correlation between individual students' scores in the three modalities was not reported, but the authors concluded that although written tests could suggest students had adequate knowledge of critical care, they may not be able to apply this knowledge in the more clinically realistic context of an OSCE or simulation. Furthermore, students scored more highly in the OSCE than the simulator, and the authors attributed this to the more demanding simulator test, which required a higher level of analysis and judgment. In OSCE stations, the clinical scenarios were broken into five separate stations, and students were given the pertinent clinical data, and guided through the diagnostic process as they were moved from one station to the next. In the patient simulator, students were required to collect, synthesise and analyse the pertinent clinical data themselves, interpret this data, and make judgements on their subsequent management. The three test modalities, written, OSCE and patient simulator, appear to test different attributes, which demonstrates the problem of attempting to validate simulation-based assessment against other forms of assessment.

Gordon et al [106] tested 23 subjects with varying clinical experience in five simulator scenarios and five oral OSCEs with equivalent content, and found good correlation between mean simulator scores and mean OSCE scores. There were significant differences in the scores between the three groups; the mean score of the ten chief residents' was higher than mean score of the seven house officers, which was higher than the mean score of the six medical students. However this study had significant limitations. The numbers were small, the authors did not report correlation between the OSCE and simulator for individual subjects in the paired OSCE and simulator assessments, and important differences may have been lost by averaging group scores. It does however add further evidence in support of construct validity of simulation-based assessment.

2.5.2 Specialist trainees/specialists

Hammond et al [56] reported scores of eight junior surgery residents, scored against a 13 item checklist in three scenarios in the patient simulator over a six month period. The three scenarios were presented in the same order, and were all scored by a single rater. Median simulator scores improved over the three scenarios, but the authors noted that in the final scenario, the resident who diagnosed and managed the arrhythmia most efficiently obtained the lowest checklist score. This study has major design limitations including small numbers, failure to randomise scenarios, and no measures to test equivalent difficulty of scenarios, but it does illustrate a potential limitation to checklist scoring, where the sum of individual items may not reflect the overall performance as judged by an expert rater.

Tsai et al [107] provide some evidence supporting reliability of simulator assessment, and certain aspects of validity. They assessed 20 paediatric residents and specialists using a child simulator in a pre-post test design where residents worked in pairs. The scenarios were of four different paediatric emergencies. Each resident was scored in one pre-course and one post-course simulation and only the lead anaesthetist was scored. The scenarios were constructed from the Advanced Paediatric Live Support course curriculum, and checklists developed for three separate scores: task specific (T), behaviour (B), and medication (M). Scoring was by the instructor, who also controlled the pace of the simulation, and by two independent raters. There was good to excellent inter-rater agreement on T, B and M scores (0.83, 0.72, and 0.71 respectively). The internal consistency (Cronbach's α) was measured on the 38 checklist items in T and the 15 checklist items in B. High levels of internal consistency were reported for B in both the pre- and post-test, and for T in the pre-test but not the post-test. Inter-case reliability was not reported. Face validity was supported by participant questionnaire feedback, and construct validity by significantly higher scores in the specialist group compared to residents. The authors used a curriculum blueprint to define the test content, and claimed content validity on this basis. Content validity also relies on sampling evenly across the blueprint, which was not demonstrated here. This study was limited by small numbers. Scoring in pairs raises the issues of the fairness of group assessment as the ability of the primary resident to correctly manage the case would be affected by the standard of help available.

The performance of a team in an emergency may have more relevance to patient outcomes than the performance of an individual. Holcomb et al [108] assessed 15 multidisciplinary military trauma teams in two scenarios using a checklist of five scored tasks and eight timed tasks related to the initial assessment of the trauma patient developed from regionally accepted standards of care. Scoring was by one of the authors. Reliability and validity of the assessment tool was not reported, but the experienced teams performed better, and the junior teams improved with training, supporting construct validity of the team assessment. As the single rater was not blinded to the experience of the participants, scoring bias cannot be discounted. From the perspective of improving patient outcomes, the ability of the health care team as a whole may be more relevant than performance of an individual, but where scores are required for certification or re-accreditation, the performance of team members may affect the scores for the individual under assessment.

2.5.3 Anaesthesia

Patient simulators were first developed for training in anaesthesia, and it is in this field that the largest body of literature on simulation and assessment lies.

In 2001, Byrne [109] reviewed the available evidence on simulation assessment in anaesthesia and concluded that more work was needed. A number of publications have advanced our understanding of simulation-based assessment in the intervening years.

Beginning with the novice anaesthetist, Forrest et al [110] described using the Delphi process to develop a valid scoring system for technical performance in emergency cases requiring rapid sequence induction. This resulted in a 91 item checklist against which the performance of six novices and seven experienced anaesthetists was scored independently by two raters. The novices were scored on a number of occasions over their 12 week training period, while the experienced anaesthetists were scored in a single case. There was good agreement between the two raters (correlation was not formally calculated), there was a significant improvement in the novices scores over time, and there was a significant difference between the scores in the novice group and the experienced group. The authors concluded simulation could be used to observe and quantify technical performance. Although the process of task analysis and the Delphi process should ensure the validity of the assessment criteria, scoring a performance against a large number of items can be problematic, and could potentially require repeated viewings of a videotape. As noted by Hammond [56], use of extensive checklists may reward thoroughness over timeliness, judgement and the ability to prioritise.

Murray et al [111] developed simulations and scoring systems to evaluate acute care skills in anaesthesia. Twenty-eight residents were scored in six simulations by six raters, using four different scoring systems: a weighted checklist, time to key actions, time to the three most important actions and a single global score which was based on time to effective management. More experienced practitioners outscored their more junior counterparts. There was high correlation between the scores using all scoring systems, but, based on generalisability theory, the global rating was most reliable. Performance varied between scenarios and the authors concluded that although simulation could be used to assess complex skills, multiple encounters would be required.

In a multi-institutional study, Schwid et al [112] recruited 99 anaesthesia residents of different grades to participate in four simulations. They used two different grading forms, two internal assessors and one external assessor, for each subject. A fourth assessor viewed all tapes for specific management errors. The grading forms had been used in a previous study [113] and consisted of either a long, weighted or a short, unweighted list of required actions, both with a complex protocol of points awarded for completion within certain times. There was a high level of inter-rater reliability (0.94-0.96) between the internal and external raters, and between the two marking formats, and internal consistency (measured between the individual scenarios) was over 0.7 (Cronbach's α). Construct validity was supported by progression of scores with increasing levels of experience. In addition, there was a modest correlation between simulator scores and results of written in-training exams and mock oral board exams, in the order of 0.44 to 0.49. They found a large number of management errors in each scenario at all grades and their description of these errors provides useful information on training needs.

Devitt et al [114-116] explored the psychometric characteristics of a simulation-based assessment in a series of studies at the Sunnybrook Health Sciences Centre in Toronto. In the first study [115], the aim of which was to determine inter-rater reliability for their assessment tool, two 60-minute long scenarios were designed, each containing five discrete and unrelated clinical problems arising during an anaesthetic (ten problems in total). Scoring for each problem was ‘no response’, ‘compensating intervention’ or ‘corrective intervention’. Investigators role-played the part of anaesthetists in these scenarios and were randomly assigned to give one of these three levels of response to each of the five problems. Three runs of the two scenarios were then videotaped, generating a total of 30 items for analysis. These were scored by two raters external to the simulations. The authors reported excellent inter-rater agreement on scores for each of the ten problems (kappa statistic of agreement = 0.96). This study, published in 1997, was one of the first attempts to examine the properties of simulation-based assessment of performance in medicine, and addressed a single psychometric property, i.e. inter-rater reliability of scoring. A major criticism of this study is that investigators were scripted to perform in the simulator at a particular level, potentially making it easier for judges to agree on the score. In an unscripted performance, there may be variation of level of performance over the problem, less clarity in completion of checklist items or otherwise, and deviation from script with alternative treatment pathways, all of which may affect inter-rater agreement.

In a subsequent study Devitt et al [116] used the same scenarios and scoring system to test internal consistency and construct validity. They defined internal consistency as a measure of the consistency in scores on the five different problems in the two scenarios. The roles of surgeon and circulating nurse were played by investigators, who also scored the participants during the simulation. Eight residents and 17 faculty members were tested in the two scenarios. Two trained anaesthetists scored the videotaped performances. Internal consistency of items within scenario one and two was 0.27 and 0.28 (Cronbach’s α) respectively. By removing two items from each of the two scenarios (missing inspiratory valve, hypotension with peritoneal traction, bradycardia with peritoneal traction and cardiac ischaemia) Cronbach’s α increased to 0.66 for the remaining six items. Overall, more experienced anaesthetists scored more highly, supporting construct validity of the assessment. However, raters also role-played the operating room team in the scenarios, and were not blinded to examinee level of experience, so that bias in scoring, or even bias in the way the scenario unfolded could have influenced the results. Inter-rater reliability was not reported, the authors assuming this was established in the previous study. However, with naïve participants instead of scripted investigators under assessment, this may not be a reasonable assumption. In an attempt to improve internal consistency, four important test items were eliminated. This may not be reasonable if test validity is to be maintained. For example, it would seem important that anaesthetists can identify a missing inspiratory valve in a patient's breathing circuit. This paper illustrates the tension between reliability and validity in test design and indicates performance varies between tasks. However, it adds support to the construct validity of performance assessment in the simulator, in that more experienced anaesthetists achieved higher test scores.

To further explore the validity of simulator assessment, Devitt et al [114] tested 33 university-based anaesthetists, 46 community-based anaesthetists, 27 final year anaesthesia residents, 37 final year medical students, and three anaesthetists identified as having practice deficiencies. The test comprised a 1.5 hour simulation containing nine discrete problems, with items and scoring system based on the two previous studies. The "surgeon" was a mannikin who could respond when necessary via a speaker from the control room to give necessary information or clarify actions. Each subject was evaluated by one of two trained raters certified in anaesthesia, who were unaware of the subject's background, but could not be blinded to their identity. Subjects rated the realism of the simulation on a ten point visual analogue scale. Two items that reduced internal consistency were eliminated (missing inspiratory valve and carbon dioxide canister leak) resulting in an internal consistency of items of 0.69 (Cronbach's α). With these items excluded, there was a significant difference in mean scores for the remaining items between all groups except university anaesthetists and final year trainees (Mean scores: university =0.53, community =0.38, trainees =0.54, medical students = 0.15) which the authors took as evidence of construct validity. The group referred for practice assessment was not included due to low numbers. All groups rated the simulation evaluation environment as realistic (7.8 on a 10 point VAS). Substituting the surgeon with a mannikin is a novel idea which could have been expected to reduce realism. In measuring complex performance in the operating room, the interaction with the surgeon seems a vital component, and eliminating this suggests the investigators were testing only limited aspects of performance.

The extent to which the reader can accept results from these three studies is doubtful. Firstly, the authors assumed that inter-rater reliability was established in the initial study, where scripted investigators role-played anaesthetists; secondly the raters also ran the scenarios and could potentially have influenced their course and degree of difficulty; and finally raters were not blinded to the experience level of the candidates.

Byrne and Jones [117] also explored construct validity, by testing 20 anaesthetists with different durations of clinical experience (<1yr, 2.5yr, 2.5-5yr, >5 yr experience) in nine brief simulations which were scored simply on time to solve the problem (slow, intermediate or fast response). There was a significant difference between the least experienced group and all other groups ($p<0.02$), but no significant difference between other groups.

It is interesting to note that both Byrne [117] and Devitt [114] could not discriminate between advanced trainees and experienced anaesthetists. In both studies the attempt to recreate the complex environment of a real operating theatre environment was limited. There is a difference in the way doctors with increasing levels of experience solve problems [7], with pattern recognition rather than analytic problem solving emerging as the dominant mode with increasing clinical experience. These simplified simulations may be appropriate tests at the early stages of training, but fail to discriminate between more advanced levels. Simple checklists may reward thoroughness over efficiency and again disadvantage the expert practitioner.

In Byrne's study [117], all groups made serious errors in diagnosis and treatment, and did not follow accepted treatment guidelines. High rates of errors in management of critical events have also been noted by Lindaker [118], where most anaesthetists made significant errors in the management of ventricular fibrillation when scored against the accepted treatment algorithm. Byrne [119] compared anaesthetists' written medical records with actual events and monitored data during a simulated case, and found major recording errors and omissions. These studies illustrate how simulation can identify widespread deficiencies in practice, rather than focussing on individual performance deficits, and identify gaps in training programmes.

Two groups of researchers have tested the performance of experienced anaesthetists working in teams in complex simulations designed to be as realistic as possible. Both evaluated inter-rater reliability for both technical performance and behaviour as the primary outcome measure. Gaba et al [120] tested 14 teams of anaesthetists and anaesthetic nurses participating in two scenarios during a series of Anaesthesia Crisis Resource Management courses. The score for technical performance awarded points for successful actions, with a number of essential actions required for a pass. Twelve behavioural criteria were developed from an instrument used to rate flight deck crew behaviours by National Aeronautics and Space Administration (NASA) with each rated on a five-point scale. The data set consisted of scores from five raters on the 12 behaviours for each team during two distinct time periods in two scenarios, and a technical score by each of three raters for each team in the two scenarios. For the technical scores, the inter-rater reliability was 0.96, but the authors concede this was artificially high due to the lack of spread of scores across the range of performance. For behavioural ratings, inter-rater agreement varied from 0.63 to 0.9, but it was noted that there was a difficulty awarding a single score for behaviour, which fluctuated over time, and there was a lack of predetermined standards against which to judge behaviour. The authors concluded that there was a probability of 0.14 that a rater would deviate by more than one point from the overall mean rating of the five judges, and a probability of 0.015 that the mean of two raters' scores would deviate by more than one point for the overall mean, suggesting two raters may produce a reproducible result for a team. There were strong correlations between the overall team behaviour and the behaviour of the lead anaesthetist, and between overall ratings and the individual behaviours. The authors concluded that it was feasible to score both technical and behavioural markers of performance in complex anaesthesia simulations, and that using two raters would limit major scoring errors. The authors also commented that prior agreement of raters on the standard of performance may have improved reliability. This study illustrates the difference between assessing simple actions in basic simulations, and attempting to assess performance more closely resembling the complexity of real life. It also raises interesting questions about the assessment of a team, rather than an individual.

Weller et al [121] evaluated inter-rater reliability when scoring overall performance of anaesthetists in simulated malignant hyperpyrexia, anaphylaxis, cardiac arrest or oxygen pipeline failure. In order to develop a valid scoring tool to assess performance in a crisis, four experienced anaesthetists agreed on a list of the required generic tasks, and from this, devised a list of observable markers against which to score performance. The draft rating form was tested and refined by viewing ten videotaped simulations, to test that each marker was observable. The resulting rating form awarded scores for medical management, behaviour, and overall performance, using a five-point rating scale with descriptors for each point on the scale. A two-hour rater training period followed, where raters individually scored the videos on the rating form, and then reconciled their scoring differences, thus arriving at an agreed standard for each level of performance. Twenty-eight videotapes from ACRM courses were rated by three primary raters and up to five additional raters (all experienced anaesthetists). There was good agreement between raters for all three components of the score, and reproducible scores could be generated by three raters (ICC 0.79 and 0.85 respectively). There was a high level of correlation between the three different scoring categories (‘medical management’, ‘behaviour’, ‘overall’), and no difference in the mean difficulty of the three scenarios. The authors commented that it was generally easy to recognise the extremes of performance, and in only one instance did one examiner award a clear fail where other examiners awarded a pass. This study explicitly described a process for developing a scoring system, and incorporated formal rater training. The use of a global score awarded by an expert judge is considered to be as reliable, and possibly more valid, than a checklist score, in the context of complex clinical performance [122]. This study found reliable results could be obtained where three raters were used to rate performance. Judging an individual within a team is potentially unfair, as other team members may be more or less helpful. Scripting helpers would increase the reproducibility of the score but decrease the validity of the team interaction.

Two other studies have scored anaesthetists' performance in emergency scenarios. Jacobsen et al [123] assessed management of simulated anaphylaxis, and found serious deficiencies in all teams, noting in particular that anaesthetists found it difficult to make a diagnosis of anaphylaxis in the simulator. The properties of the scoring system and the correlation between treatment sequence and ACRM scores were not reported, making it difficult to interpret the reproducibility of their findings. In a second study, the group observed anaesthetists' management of malignant hyperthermia [124]. Again the properties of the scoring system were not analysed, but the authors reported all teams made the correct diagnosis and all gave dantrolene, but there was some variation in other items. In particular they found that many teams failed to hyperventilate the patient. These two studies identify common errors in treatment, and, like the study by Schwid et al [112], should inform the development of an emergency care curriculum as well as identify the needs of individual practitioners.

Most studies described so far have evaluated aspects of validity and reliability of simulation based assessment. Additional important attributes of an assessment process are its acceptability and the impact on student learning. There is some evidence on the former. In 1993, Riley [125] conducted a postal survey of anaesthetists in Australia and the USA and found that the 183 responding anaesthetists were generally supportive of simulation for training but not of the use of simulation in certification and recertification. In 1994, Kurrek et al [101] conducted a postal survey of 150 Canadian anaesthetists and found low support for compulsory use of simulation in recertification. However the response rate was only 39%, and 19% of respondents had never heard of simulation.

Holzman [75] et al, in 1995, surveyed 75 anaesthetists and nurse anaesthetists and found the majority thought simulation would be useful in certification, but significant numbers disagreed.

2.6 Summary

These studies shed some light on simulation-based assessment in a number of contexts. An assessment tool must be shown to be reliable, valid, feasible, acceptable to examinees and have a positive effect on learning.

Many studies have evaluated inter-rater reliability, and in general have found reasonable agreement between raters. It is likely that agreement between raters will be greater when the simulation is brief and straightforward, for example, assessing airway management by medical students in contrast to assessing a complex anaesthetic emergency scenario managed by a specialist and their team.

Most studies have used checklists, which may be more appropriate at the medical undergraduate level than in assessing the more complex performance of experienced practitioners. However a number of studies have indicated reasonable agreement between checklist scores and global scores awarded by expert raters. Internal consistency of checklist items, or between different problems on the simulator, has generally been shown to be low, and to obtain an adequate level, items have been removed. This is problematic where the checklist item represents an important attribute. Only in more recent studies has the variability of performance between cases been studied, with evidence emerging on the need to include a large number of cases to obtain a reliable result. The extent to which this extends to the higher levels of performance, where it could be expected that some of the attributes of crisis management are generic, has not been examined.

Many studies have examined construct validity, demonstrating on the whole that scores are higher in groups where a higher level of performance is expected. Several studies however, fail to demonstrate a difference between senior residents and specialists. It may be that there is no difference, or that the subtle cues that prompt an expert practitioner to recognise a particular pattern of presentation may be missing in the simulator, forcing more reliance on analytical thinking. Alternatively, the simulator technology or ability to engage in the role play may discriminate against older practitioners.

Another important aspect of an assessment is content validity, i.e. ensuring that the assessment tests across the curriculum blueprint. A number of studies did address this issue in their design. Face validity was demonstrated in a few studies, and is also supported by reports of course participants attesting to the realism of the simulations.

Examinees will generally find an assessment acceptable if it is seen to be fair, and appears to be a valid measure of important attributes. A few studies report on this for undergraduates, but the only evidence at the post-graduate level is from postal surveys conducted some years ago, at a time when many survey recipients had limited or no experience in simulation.

None of these studies report on the effect of simulation-based assessment on student learning. It would be expected that an assessment that focused on clinical skills and patient management would drive students to spend more time learning in the clinical setting rather than reading in the library, but this remains to be tested.

This review has mapped the current evidence on patient simulators, and identified a number of gaps in our knowledge. The following chapters will describe a series of studies designed to fill some of these gaps and add to our understanding of simulation in medical education.

Chapter 3 Continuing medical education: simulation and change in practice

3.1 Introduction

Doctors are expected to keep up to date and remain competent practitioners through continuing medical education (CME). Indeed, vocational registration with the New Zealand Medical Council is dependent on participation in an approved Maintenance of Professional Standards (MOPS) programme put forward by the relevant Medical College. For many specialists, this can be satisfied by attendance at conferences, refresher courses or workshops. However, there is substantial evidence that, despite their popularity with doctors, many approved CME activities do not ensure continuing competence, or currency of practice.

In order to determine the place of simulation in continuing education for doctors, this chapter begins with a review of the evidence on the effectiveness of CME activities [93]. Three studies evaluating the effectiveness of simulation in the context of continuing medical education will then be described.

3.2 Searching for evidence on educational interventions: does continuing medical education work [93]?

Systematic reviews of the educational literature suggest that a good deal of CME achieves little in terms of changing physician behaviour or improving patient outcome [126, 127], and more effective educational strategies, potentially incorporating simulation-based activities, could be used. Outcome measures used to determine CME effectiveness include improved patient outcomes, observed changes in physician practice, improved test scores, or self-reported changes in confidence or practice.

Improved patient outcome is the over-riding goal of medical education, but is often difficult to demonstrate. Where improved patient outcomes have in fact been linked causally to a specific educational intervention, the outcome has been very specific, and closely coupled with the intervention. Examples of this are lower HbIAc levels in a group of diabetic patients whose doctors participated in an intensive programme on diabetic management, and lower rates of back surgery in randomly selected communities participating in a programme involving spine surgeons, community physicians, patients and hospital administrators [128, 129]. Unless the intervention is closely linked to an explicit and measurable patient outcome, there are generally too many factors between the educational intervention and the patient outcome to attribute causation. College CME activities tend to be more general in their scope, and rarely defined in terms of measurable patient outcomes. Furthermore, the most educationally sound intervention may fail to show an improved patient outcome due to factors beyond the control of the physician or educator. A patient may have been correctly prescribed a new, evidence-based treatment, but chosen not to comply, or reacted idiosyncratically to the medication. As an example, in a meta-analysis of interventions aimed at improving physicians' management of hypertension, there was no change in patients' blood pressures despite improved patient follow-up, and improved patterns of prescribing [130]. In addition, it may be difficult to demonstrate an effect where an intervention is aimed at a relatively rare event. For example, Combes et al [131] developed guidelines for management of the unanticipated difficult airway and introduced them into practice in a department of anaesthesia. However, in an 18-month prospective trial of over 11,000 intubations, no cases of failure to intubate or ventilate an anaesthetised patient occurred. It would clearly be difficult to demonstrate improved patient outcomes for this intervention.

Using changes in physician practice as an outcome measure avoids unpredictable patient factors [130]. Nyman et al [128] used this approach in a before/after study evaluating a multi-component intervention to improve diabetic patient management, and reported positive changes in the rate of urine testing, foot examinations and dilated eye examinations by participating general practitioners.

In contrast, numerous studies have reported no change in doctors' behaviour after lectures, receipt of written evidence-based management guidelines, or participation in traditional CME activities [126, 128, 132, 133]. Referral practice with stroke victims was found not to change following attendance at a lecture on guidelines for management of stroke [134]. Evidence-based guidelines on management of acute myocardial infarction have not been adopted universally, despite numerous publications and presentations [135, 136].

A frequently used method to evaluate an educational intervention is to compare pre and post-intervention test scores of participants [134, 137, 138]. In this way it is possible to show that physicians often know more about a topic after they have spent time reading, listening to a presentation, or engaging in a small group discussion, but it is a big jump to assume this increase in knowledge will translate to a change in physician practice and ultimately improved patient outcomes. However, with appropriate study design, pre and post intervention test scores may be a useful measure. For example, in an evaluation of a programme in adolescent health offered to a self-selected group of general practitioners, consultations with standardised adolescent patients were videotaped and scored for quality over the course of the intervention and again over the following 12 months. Large improvements were demonstrated compared with a control group of non-participating doctors [136].

CME participant opinion, including self-reported changes in practice are an alternative method to evaluate CME activities [138-141]. For example, participants in a practice-based CME programme delivered by video-conferencing reported specific practice changes they had made as a result [138]. GPs receiving a self-learning package on diagnosis and management of STDs reported increased knowledge and skills and changes to their practice [139]. Interviews of six general practitioners attending an Annual Refresher Course reported that traditional CME programmes had changed their practice, were able to give concrete examples of these changes, and were sceptical of the concept that practice changes did not result from traditional CME activities [140]. De Villiers et al [141] demonstrated support for a shift from the traditional update lecture series to a competency based model for continuing medical education among general practitioners. Participants reported that the small group techniques could identify specific learning needs and allowed in depth exploration of topics, resulting in self-reported improvements in patient care and clinical skills.

Evidence from the literature, and theories of how doctors learn, offer some guidance on what types of CME activities should be effective. The systematic reviews conclude that traditional medical conferences, printed educational materials and didactic lectures have only weak effects in terms of change in physician practice [126, 127, 142]. Simply providing people with written information is not a reliable way to change clinical practice [135]. On the other hand, well-constructed interactive sessions can lead to change in physician practice [126, 127]. Clinical audit and providing feedback to clinicians on their professional practice have also been shown to be effective, though in the meta-analysis by Jamtvedt et al, using the limited evidence available, the effects were small [143]. At a departmental level, local, highly respected doctors can play an important role as opinion leaders, capable of influencing the practice of their workplace colleagues. These people could be a valuable resource for championing changes in workplace practice [144]. Meetings that provide an opportunity for small group interactions, and learning centred around workplace practice seem to be effective [145]. Other innovations that show some evidence of being effective include practice-based initiatives providing structured time to work with colleagues [146], and portfolio-based learning programmes such as those available to Canadian physicians [147].

In order for new material to be learnt, it must be processed and integrated into existing frameworks. An experience, be it a concrete clinical experience or a more abstract exposure to new ideas, must be acted upon in some way in order for it to be incorporated into long term memory and change behaviour. Kolb [148] suggests that this is done through reflection or active experimentation with the new material. CME activities which appear to be effective are those which present learners with a new experience or concept, and require the learner to actively engage with this experience or concept to create from it their own meaning.

With this background of empirical and theoretical evidence on effective CME, three studies were undertaken to determine the position of simulation in continuing medical education. The first study maps the overall picture of continuing education for New Zealand anaesthetists, and where anaesthetists see simulation in this picture [149], and the second study focuses on the ability of simulation courses to change the practice of anaesthetists [150]. The third study looks at extending simulation-based training beyond the hospital-based specialties to the management of medical emergencies in primary care.

3.3 Simulation in the CME continuum: A survey of current practices in Continuing Education in New Zealand Anaesthetists [149]

The aim of this survey was to define the current pattern of participation by New Zealand anaesthetists in CME, to evaluate which activities were effective in changing practice, or considered useful, and to identify the motivators and barriers to participation in CME activities. The hypothesis was that anaesthetists' perceptions of effective CME activities would reflect the findings in the systematic reviews of the CME literature.

3.3.1 Method

In October 2002, following a pilot of the questionnaire by ten anaesthetists, an anonymous survey was sent to all New Zealand anaesthetists on the mailing list of the Committee for Continuing Education in New Zealand (CECANZ). This included all New Zealand Fellows of the Australian and New Zealand College of Anaesthetists (ANZCA) or members of the New Zealand Society of Anaesthetists. The analysis was restricted to responses from those identifying themselves as specialists. The returned surveys were separated from the coded envelopes on opening, to preserve anonymity. A second survey was mailed out one month after the first to those anaesthetists who had not yet responded.

The survey contained 17 questions. Seven were open questions inviting written responses, while the remainder required a choice of options or rating on a scale of 1-5 (Appendix 1).

Anaesthetists were asked if they worked in private, public or other types of hospitals, and the number of colleagues in their workplace as an indication of practice location, where 11 or more colleagues indicated a large centre and less than 11, a small centre.

Anaesthetists recorded participation in specific CME activities and rated the usefulness of these different activities in terms of changing their clinical practice on a scale of 1 to 5. They were asked to identify changes they had made to their practice as a result of this participation, and what activity was responsible for this change. Written responses were sought on preferred types of sessions at conferences, and the factors that increased motivation to participate in CME. Anaesthetists nominated barriers to participation from a list of options. Finally, suggestions for a future CME programme were invited.

To compare the ratings for usefulness of different activities, a mixed model two-way analysis of variance was used, with activity rated as a fixed effect and individual ratings as a random effect. An adjustment was made for multiple comparisons. Chi squared test was used to determine the significance of differences in proportions between groups. Written comments were grouped into themes and numbers of like responses expressed as percentages of the total number of responses to each question. At no point in the survey were anaesthetists prompted to specifically consider simulation in their responses.

3.3.2 Results

A total of 311 forms were returned from specialists. Of these specialists, 16 were retired and three practised outside anaesthesia. These were excluded, leaving 292 forms with sufficient information to be included in the analysis. This represents 74% of the 393 vocationally registered anaesthetists with annual practicing certificates as of December 2002.

Of these 292 specialists, 273 worked in public hospitals, 197 in private hospitals and one in a military hospital. Nineteen of these anaesthetists (6.5%) worked solely in private practice, and 248 (85%) worked in hospitals accredited for training by the Australian and New Zealand College of Anaesthetists. Of the 286 respondents to the question, 85% worked in large and 15% in small centres. In terms of overall satisfaction with their current CME activities, 80.5% were satisfied and 19.5% were dissatisfied (278 respondents).

Fifty-five percent of anaesthetists attended more than ten department meetings per year, 31% attended between six and ten per year and 14% attended less than six meetings. Anaesthetists in private practice attended significantly fewer department meetings ($p<0.01$), with 36.8% of those working solely in private practice attending less than six meetings per year, compared to 14% in the group as a whole (280 responses to question). There was no significant difference in the number of meetings attended by anaesthetists in large and small centres.

Responses to the question, “Which of the following CME activities have you undertaken in the preceding 2 years?” are shown in Table 3. The “HELP module” was a four-monthly publication consisting of a booklet of multiple choice questions around a particular theme, and an accompanying answer booklet, which was sent to all New Zealand anaesthetists on the CECANZ mailing list from 1985 until 2003. Single theme meetings are one day updates on a clinically relevant topic attracting an audience from around New Zealand.

Ratings for usefulness of CME activities are shown in Table 4. There were significant differences between the perceived usefulness of these activities, with simulation courses and skills workshops rating significantly more highly than single theme meetings, conferences and HELP modules (p<.0001). Single theme meetings and department meetings were considered the most useful meetings, rating significantly higher than the CECANZ Annual Scientific Meeting (p<.005). The HELP modules were rated significantly lower than all other listed choices (p<.005).

Table 3: The number of anaesthetists reporting involvement in each of a list of CME activities in the preceding 2 years. (%of the 292 responses)

CME Activity	**number (%)**
CECANZ ASM	186 (64)
HELP modules	182 (62)
Single theme meeting (NZ)	180 (62)
Simulation crisis management course	154 (53)
ANZCA ASM*	144 (49)
International conferences	130 (45)
Skills Workshop	107 (37)
Other Australian Meetings	77 (26)
ASA* ASM	61 (21)
EMST* course	41 (37)
Other	64 (22)

*ASM=Annual Scientific Meeting.

*EMST: Early Management of Severe Trauma.

*ASA=Australian Society of Anaesthetists.

Table 4: The ranking of different CME activities

		Mean rating	Number rating activity
Most useful	Simulation Centre Crisis Management Course	1.70	228
	Skills workshops	1.80	224
	EMST	2.00	142
	Single Theme Meetings	2.09	251
	Department meetings	2.20	292
	Australian conferences	2.26	240
	International conferences	2.28	237
	CECANZ ASM	2.44	255
Least useful	HELP modules	2.81	259

Ratings on a scale of 1-5 where 1=very useful, 5 =not at all useful. Number rating activity is the total number of anaesthetists rating that activity.

There were 258 written responses to the open question, “When attending a conference, what types of sessions do you find most useful?” Many offered more than one answer. Interactive formats were found most useful by 72%, with 48% wanting workshops, sessions promoting discussion (8.5%) small group discussion (8%) debates (7%) while 25.5% found lectures most useful. Relevance to clinical practice and updates were preferred by 38%, input from experts by 11% and focus on a single theme by 6.5%.

Of the 261 responding to the question, 93% said they had changed their practice as a result of CME in the preceding two years and 70% described a specific change they had made. Five percent had not changed their practice and 2% had confirmed their practice was acceptable. Some anaesthetists reported more than one change.

Specific changes fell into a number of themes including revised indications for epidural anaesthesia and peripheral nerve blocks, use of peri-operative beta blockade in patients at risk of cardiac ischaemia, multimodal anti-emetic strategy and more aggressive treatment of post-operative pain. Changes to airway management included new indications for the use of the laryngeal mask airway and its variants, use of airway algorithms, more thorough airway assessment, awake fibreoptic intubation and more confidence with surgical airways. Behavioural changes included change in leadership style, task delegation and improved communication during a crisis.

Changes in quality assurance activities included audit of practice, a move to evidence based practice, development of protocols, a change in approach to consent, safer measures such as "methods to reduce drug errors" and a change in approach to pre-assessment, typified by the comment "I now allocate more time for preoperative assessment" (Table 5).

Table 5: Changes to practice reported by 182 anaesthetists (open question)

Reported changes in practice	number (%)
Change in use of regional techniques	50 (27)
Changes in Airway management	40 (22)
Peri-operative management of patients with CVS risk	35 (19)
Change in management of PONV	30 (16)
Change in communication or behaviour in a crisis	33 (18)
Quality assurance activity	26 (14)
Change in acute pain management	22 (12)
Changed approach to pre-assessment	10 (5)
Other	58 (32)

Two hundred and twenty-four anaesthetists identified one or more activities that had led to this change. These were: conference or course 51%, departmental meeting 23%, simulation centre course in crisis management 15%, reading 14%, skills workshop 9%, talking to colleagues 4.5%%, HELP module 2.5%, hospital attachment 2.5%, EMST or similar course 2%. The relative effect of the activities in changing behaviour cannot be assessed due to the different times and participation rates for each.

Two hundred and seventy five anaesthetists described their motivation to participate in CME (Table 6). The requirement for Maintenance of Professional Standards points (MOPS) emerged as the dominant theme, although only ten anaesthetists put this as the only motivation. Keeping up to date and ensuring patient safety were also frequently reported, as were the quest for knowledge, self-esteem and respect of colleagues. Location of a meeting did not emerge as a strong motivating force.

Table 6: Motivation for participation in CME. (275 written responses to an open question, more than one motivating factor suggested by some respondents)

Motivating factor	number (%)
Need for MOPS points.	101 (37)
Keeping up with new developments, updating practices	98 (36)
Curiosity, interest, desire to learn	88 (32)
Self-esteem, self-development, professional pride, peer pressure, competition	60 (22)
Improving patient safety and outcome	57 (21)
Ensuring high standard of practice, self-audit of practice	27 (10)
Talking to colleagues	23 (8)
Medico-legal concerns	9 (3)
Opportunity to travel, interesting location, break from work	6 (2)

Anaesthetists marked on average, 1.6 major impediments each to their participation in CME activities from the list of options (Table 7). Anaesthetists solely in private practice on average identified 2.6 impediments each. Getting time off work was significantly more of a problem in private than in public practice (53% versus 28% respectively, $p<0.02$). However, anaesthetists working in public found other work commitments more of a problem than those working only in private, ($p<0.05$). Those in small centres found distance more of an impediment than those in large centres ($p<0.02$). Reporting of other impediments was not significantly different between groups.

Table 7: Impediments to participation in CME. (% of 292 responses. 17 did not report any impediment)

Impediment	number (%)
Other work commitments	154 (53)
Distance	76 (26)
Getting time off	84 (29)
Funding	57 (19)
Lack of relevant content in meetings	46 (16)
Other	55 (19)
No impediments	17 (6)

There were 167 suggestions for future programmes, which reflected previous indications of preferred or useful activities, and included many requests for specific topics.

3.3.3 Discussion

New Zealand anaesthetists are active participants in formal CME. Most seem satisfied with their CME and report changes in their practice as a result of their participation.

In contrast to recent reviews in the Cochrane database [126, 133] examining outcomes of CME interventions, respondents to this survey described many very specific changes to practice as a result of a variety of CME activities.

In Davis' meta-analysis [126] there was some evidence that activities where participants were actively engaged and had the opportunity to practice skills could cause a change in professional practice and, on occasion, health care outcomes but conferences were unlikely to produce a measurable change. The high value placed on simulation and skill workshops by survey respondents supports this conclusion. One third of respondents identified change in their behaviour in a crisis, or in the way they communicated to their operating room team, and over half identified patient safety as a key motivator in continuing education.

However, a range of other activities were also found to be beneficial. Conferences appeared to be effective, and were reported as the activity responsible for a change in practice on 114 occasions. The specific conference activity responsible was not identified but this does suggest we should not abandon conferences. O'Brien [144] suggested local opinion leaders could be a powerful tool for change by sanctioning the introduction of new practices, which may be why, at conferences, visiting speakers and lectures were found useful. Departmental meetings were rated highly and reported to produce change in practice, and in addition there were clear preferences for interactive, relevant meetings. On theoretical grounds, department meetings are likely to be effective as they promote small group discussions and learning based on clinical practice [145].

CME embraces the concept of life-long learning, a continual process of reflection and self-assessment. This study was limited to traditional CME in the form of externally organised activities. A learner-centred approach begins with the learner identifying the need for change. The CME programme should facilitate both the process of identifying the need and ensuring the means for change is available.

This study provides an overview of CME in New Zealand, and provides a perspective on the relative value of simulation-based courses in comparison to other activities. Unprompted, anaesthetists identified simulation as being the most effective form of CME in terms of producing a change in their practice. This is strong and independent evidence in support of simulation-based education.

The following study explored further what anaesthetists found valuable in simulation-based courses in their continuing medical education.

3.4 Change in practice following simulation training in crisis management: a survey [150]

The purpose of this study was to investigate the long-term effects of a simulation based course in anaesthesia crisis management on subsequent clinical practice. Simulation-based training in crisis management is a popular form of CME for anaesthetists in New Zealand, as identified in the previous study. Over a three year period from the beginning of 2000 until the end of 2002, 199 specialist anaesthetists attended crisis management courses at the National Patient Simulation Centre in Wellington, New Zealand. This represents almost half of the anaesthesia specialists in the country.

As previously discussed, evaluation of the outcomes of educational interventions can be problematic, and finding evidence for improved outcomes following training in crisis management presents particular difficulties. Because of the good record for safety in anaesthesia there are few adverse outcomes, so to demonstrate a reduction in adverse outcomes from any intervention would require a sample size of very large proportions.

An alternative approach would be to look for improved clinician performance, and in the case of crisis management training, it would be desirable to demonstrate improved management of crises in the operating theatre. Problems with scheduling a standardised, repeatable crisis, and the ethical issues of observation without intervention, limit the reliability and feasibility of this option. It seems unlikely that a randomised controlled trial will be applicable in this context.

A proxy measure of training effectiveness could be improved performance in the simulator. This would be a more reliable and feasible proposition [121], but rests on the assumption that performance in the simulator predicts performance in clinical practice. Byrne and Greave [109] reviewed the evidence on assessment of performance during anaesthetic simulations, and suggested the validity of simulator performance as a measure of real life performance remains to be tested.

As an alternative approach, we looked for self-reported change, and investigated anaesthetists' own perceptions of changes in their clinical practice as a result of attending a course on effective management of anaesthetic crises at the National Patient Simulation Centre.

3.4.1 Methods

In February 2002, a questionnaire was posted to the 96 anaesthetists who had attended a one-day Anaesthesia Crisis Management (ACM) course at the National Patient Simulation centre in 2001 (3-12 months prior to the mail-out). The ACM courses were limited to a maximum of six participants, and over the course of the day, groups of two or three anaesthetists took it in turns to manage a series of four to six clinical crisis scenarios played out on the high fidelity simulator. Anaesthetists not involved in a scenario watched by video relay in an adjacent room. Both participants and observers took part in a facilitated discussion after each scenario, during which time they reflected on the events and behaviours that occurred, and developed plans for dealing with similar events or interactions in the future. To aid reflection, these discussions incorporated video replay of illustrative segments of the scenario.

The questionnaire asked participants to rate the relevance of the course to clinical practice (using a scale of 1-5 where 1 = not relevant and 5 = highly relevant, and to rate the extent to which the course had increased their confidence in crisis management (using a scale of 1-5 where 1 = decreased and 5 = markedly increased). Written comments were invited on changes in confidence in crisis management, changes to routine clinical practice as a result of the course and key points learnt from the course. Participants were asked to describe their involvement in actual clinical crises in their subsequent practice and any perceived changes to the management of these crises as a result of the course. Participants were also asked if they intended to return for follow-up courses. Responses were anonymous. There was a second mailing in April 2002 (5-14 months after attending the course).

Written responses were coded and grouped into themes. The coding, identification of themes and the categorisation of responses into these themes was undertaken by two specialist anaesthetists and agreement reached on the analysis.

3.4.2 Results

Sixty-six responses were received, (42 from the initial mailing and a further 24 from the second mailing), a response rate 69.5%. One survey was returned unopened. From course registration information, it was known that of the 96 course participants, 14 (15%) were anaesthetic registrars in training, and 82 (85%) were specialist anaesthetists. As the questionnaire responses were anonymous, it was not possible to identify the composition of the respondents in terms of their level of experience.

Median (interquartile range [range]) and score for relevance to clinical practice was 5 (4-5 [2-5]) and score for increased confidence in crisis management was 4 (4 [2-5]).

Thirty-six respondents (55%) offered reasons why they felt more confident (Table 8). Participants felt they had “better team organisation skills”. Senior anaesthetists valued the opportunity to practice for crises: “Usually situations are confronted by junior staff first, so (my) knowledge gets rusty." There was evidence of change in attitudes, for example, the course "gives one the confidence to get help without feeling you've failed for not sorting it out yourself."

Table 8: Reasons for increased confidence in ability to manage a crisis following simulation course, given by 36 respondents. Values are number (%).

Opportunity to practise	12 (33)
Better teamwork skills and delegation of tasks	11 (31)
Greater control and less apprehensiveness	8 (22)
Better communication	4 (11)
Greater ability to stand back, review and re-evaluate	4 (11)
Greater willingness to seek help	4 (11)

Thirty-six respondents (55%) described changes they had made to their routine practice following the course (Table 9). Changes in the way they communicated with the rest of the operating room team emerged as a major theme, with references to "more formalised handovers for breaks", increased sharing of problems and management plans with the rest of the operating room team and more explicit definition of roles of team members as exemplified by the comment "I have been more conscious of clearly delegating tasks rather than just expecting it will happen." Planning for possible adverse events included insisting on skilled assistance, rehearsing treatment algorithms and calling for help earlier. Changes to problem solving strategies included regular global assessment, more awareness of the possibility of fixation errors and thinking out loud or using protocols. It was interesting to note that three respondents (8%) were taking actions to improve the communication, teamwork and technical skills of their operating room colleagues and junior staff.

Table 9: Changes to routine practice following simulation course given by 36 respondents. Values are number (%).

Communication with colleagues	12 (33)
Working with a team	11 (31)
Planning for adverse events	16 (44)
Problem-solving strategies	5 (14)
Training colleagues in crisis management	3 (8)

Key points learnt from the course are shown in Table 10. Participants cited “The value of nominated, clear leadership” and being “more proactive”. Clarity of communication, the advantage of knowing staff names, and the importance of a clear and concise description of the case to colleagues were mentioned. Consideration of problem recognition and problem solving strategies during critical events was an important issue for many participants. Comments included the importance of not making assumptions, of a systematic approach, of using colleagues to help problem solve and the need to keep an open mind, as typified by the comment: “How easy it is to get into a closed mindset. [I see] the importance of re-evaluating, even when you’ve made up your mind.”

Forty respondents (61%) had experienced one or more critical event in their clinical work following the course, including six cardiac arrests, six major haemorrhages, seven anaphylactic reactions, a number of obstetric emergencies (including two cardiac arrests secondary to air or amniotic fluid embolus) and seven airway emergencies. Twenty-eight of these 40 respondents (70%) felt their management of the crisis was improved as a result of participation in the simulation course, 10 were unsure and 2 reported no change. Fifty-eight respondents (88%) indicated they would return for a further Anaesthesia Crisis Management course, six were unsure and one did not plan to as s/he was about to retire.

Table 10: Key learning points obtained from simulation course, given by 62 respondents. Values are number (%).

Communication skills	23 (37)
Problem-solving strategies:	25 (40)
Systematic approach to a problem	3 (5)
Global assessment of the situation	3 (5)
Re-evaluation to avoid error	7 (11)
Not making assumptions	3 (5)
Value of verbalising the problem	3 (5)
Sharing the problem with team members	6 (10)
Teamwork in a crisis	20 (32)
Importance of delegating tasks	16 (26)
Leadership issues:	17 (27)
Need for a leader	9 (15)
Ability to organise the team	6 (7)
Prioritising tasks	2 (3)
Calling for help from colleagues	7 (11)
Anticipating or planning ahead	4 (6)

3.4.3 Discussion

The fundamental purpose of CME is to improve outcomes for patients by improving physician performance. Evidence for the effectiveness of different forms of CME has been systematically reviewed over recent years [126, 133] and the evidence suggests that many activities have little effect on practice. Interactive workshops and opportunities to practice skills are more likely to result in change in practice. The crisis management course fits this model and should be effective in changing practice [145].

Simulation-based training in anaesthesia crisis management was introduced over a decade ago [64] but there are limited publications on the effectiveness of this training. Data from questionnaires, completed at the end of the course suggest that participants see anaesthesia crisis management courses as relevant to clinical practice, with high levels of participant satisfaction in the training [74]. There is some evidence that simulator training improves subsequent performance in a simulated crisis [73]. However, there is little information on long-term changes to anaesthetists' practice following simulation-based crisis management training.

This survey provides evidence that the anaesthesia crisis management course at the Wellington Simulation Centre resulted in a long term change in practice. The changes reported include the ability to work collaboratively with colleagues, to be a good communicator and to be able to lead a team effectively. Explicit training in these attributes is limited in current CME programmes in anaesthesia. Of interest is the change respondents perceived to their routine practice as a result of the course. Improved interactions between team members and better strategies for problem-solving may be effective in preventing, as well as managing, crises.

A surprisingly high proportion of respondents were involved in one or more major adverse events subsequent to the course. If one assumes that the respondents were not markedly different from the general population of anaesthetists, there appears to be a high likelihood that an anaesthetist will encounter a clinical emergency each year. This is strong evidence in support of the need for continuing education to ensure anaesthetists can manage crises effectively. Of the many adverse events described, only three were clinically similar to any of the six events simulated during the simulation courses. However, 70% of respondents who had experienced an adverse event felt they managed it better as a result of their attendance at the course, suggesting that the skills acquired in the course could be generalised to the management of a wide range of crises.

A response rate of 69% is not uncommon for a postal survey but nothing can be said about the 31% who did not respond. Those who gained most from the course, or who subsequently experienced a crisis, may have been more likely to reply to the survey. In addition, this study is limited to anaesthetists' self-perceptions of change in practice. It would be desirable to have a more objective outcome measure. A clinician's own perceptions of his or her abilities may be inaccurate, but the specific examples of modifications to practice described by the respondents lend weight to the concept that there has indeed been a change in behaviour.

3.5 Simulation training for medical emergencies in general practice [151]

The two preceding studies focussed on management of emergencies in anaesthesia. The following section describes a pilot course developed for general practitioners for managing the critically ill patient in the context of primary care, and the evaluation of this course by participants.

3.5.1 Context and setting

Medical emergencies in suburban general practice occur rarely, making it difficult for general practitioners to develop and maintain relevant expertise. The use of a patient simulator and appropriate vignettes could recreate these events in a context relevant to general practitioners, allowing rehearsal and planning. In addition to updating and practising treatment protocols, doctors could identify issues in their ability to respond to emergencies in their practice setting and devise strategies to deal more effectively with these events.

3.5.2 Why the idea was necessary

General practitioner colleagues expressed anxiety at the possibility of patients with acute, life-threatening events presenting to their practice. The equipment and support available is considerably less in a suburban general practice than in the hospital setting. Management decisions often relate to expeditious transfer of care, and deciding on the risk versus benefit of embarking on potentially dangerous interventions in practice rooms. My general practice colleagues suggested the range of acute medical conditions, and the management issues relevant to general practice were not adequately addressed by the Advanced Cardiac Life Support (ACLS) courses they were required to attend to maintain accreditation with the New Zealand Medical Council.

3.5.3 What was done

We obtained ethics committee approval. Staff at the National Patient Simulation Centre (myself and the director of the centre), two academic general practitioners at the Wellington School of Medicine and two community general practitioners in Wellington, New Zealand, together developed a one-day simulation course based on six clinical scenarios the general practitioners agreed were of particular concern to them. These were anaphylaxis, congestive cardiac failure, myocardial infarction, cardiac arrest, tension pneumothorax and life-threatening arrhythmia. We used a Laerdal Simman patient simulator, employed a professional actor as the “relative”, and recruited a nurse familiar with working in the simulation centre to work with participants during the scenarios. Both general practitioners and simulation centre staff instructed on the course.

Eight general practitioners attended the course as participants. Each took part in a simulated emergency, while others watched through video relay. A debrief and clinical update followed each simulation. All participants were asked to complete a post-course evaluation which included both limited choice and open questions.

3.5.4 Evaluation of the results

All participants completed the evaluation form at the end of the day and overall responses were positive. All agreed the workshop was pitched appropriately, and six felt their emergency skills had improved (two were unsure). Seven would recommend the course to colleagues (one unsure). The median rating for usefulness was 2, (range 1-3, 1=very useful and 5=not useful). Median score for realism was 2 (range 2-5, where 1=very realistic, 5=not realistic). The most realistic aspects of the scenarios were reported as the interactions with nurse and relative, the general lack of equipment, the unfamiliarity with equipment and feeling stressed. Least realistic was the lack of visual cues from the simulator.

Key learning points identified in answers to open questions were decision making, teamwork and communication, practical realities of handling emergencies in general practice and resuscitation skills update. All participants had recently attended an ACLS course. When asked to compare this with the simulator course, they found the simulator course more relevant to the general practice context, included a wider range of clinically relevant situations, focused more on teamwork and communication, was more practical, focused less on algorithms, was more like real life, and more memorable.

Seven general practitioners identified changes they intended to make at their practice. These included organising on-site training sessions with practice staff, sorting out emergency equipment and protocols, and improving staff interactions.

Suggestions for improvement were to run the course in with the simulator set up in a real general practice, to seek accreditation of the course for ACLS (which could be achieved with only minor additions to course content), and to develop a learning module on emergency response to be used in general practices in other suburban contexts.

3.5.5 Summary

This small pilot study indicates there is the potential to apply lessons learnt in hospital-based simulation courses in crisis management to the primary care context, but modification is required. A number of lessons were learnt in terms of appropriate course design. The first was around creating realistic simulations for general practice. Some of the more realistic aspects of simulations in anaesthesia are the interface with the monitors displaying the patient's vital signs, and the interactions with the operating room team. However, in general practice, there are few monitors available, more reliance is placed on history and physical examination, and as general practitioners usually work on their own, the interaction with the patient is paramount. Alternative approaches are required to enhance realism. Secondly, the input into the course curriculum from general practitioner colleagues was vital to ensure relevance to the participants. Issues for primary care physicians were very different to those in hospital-based practice both in the type of clinical events that cause concern, and the range of responses to these events that are relevant in general practice. In addition, although the participants in this course were enthusiastic, we identified a number of barriers to widespread participation of general practitioners in this form of training. On the whole, general practitioners are not provided with employer funding or protected leave for continuing medical education activities. This places considerable limitations on the length and scheduling of courses and the level of course fees that can be charged. The intention had been to offer courses to 20 general practitioners, but this proved to be problematic, despite a grant enabling the pilot course to be offered at minimal cost. Prior ACLS accreditation of the course may have improved recruitment.

This study did however support the need for improved responses to uncommon but life-threatening emergencies in general practice, and suggested that the maximum benefit would be gained from taking the simulator and instructors out to the practice rooms, and involving all the staff in the practice including nurses and receptionists. The simulations could be used to test the adequacy of facilities, protocols, and the responses of the practice staff to emergencies in their own environment. The Laerdal simulator used in this study is relatively transportable, making this a feasible option.

3.6 Conclusion

A review of CME interventions suggests that many traditional forms are ineffective and unlikely to ensure ongoing competence of participating doctors. Where CME activities are targeted to specific outcomes, closely linked to workplace practice, and engage participants in reflection and active processing of material, change in physician practice is more likely. Educational activities incorporating patient simulators enable active engagement of participants in a realistic clinical experience and facilitate reflection on the experience and as such, are likely to have a powerful impact. Evidence from these evaluation studies supports this proposition, where simulation courses were ranked as the most effective CME activity undertaken, course evaluations are positive and many participants in simulation-based courses identified specific changes they had subsequently made to their practice.

Simulation-based training for crisis management is effective, and simulation-based courses should be used increasingly in programmes for continuing medical education. The evidence provided here is not limited to anaesthesia practice, but would apply widely to the continuing educational needs of all physicians managing acutely ill patients.

Chapter 4 An argument for including simulation in the medical undergraduate curriculum in critical care

4.1 Introduction

One of the expected outcomes of undergraduate medical education is the ability to recognise and manage critically ill patients. However, there is increasing evidence that a problem exists, and poor care may be contributing to excess morbidity and mortality in hospital inpatients.

McQuillan et al [152] investigated the quality of inpatient care in the hours leading up to unanticipated admission to the Intensive Care Unit (ICU), and identified major deficiencies. Over half of the admissions were considered to have had substandard care, where criteria for care included appropriate oxygen therapy, management of the airway, breathing and circulation and optimal use of monitoring. The main contributing factors were lack of knowledge, failure to appreciate urgency, lack of experience, failure to seek advice and lack of supervision. Their recommendations included organisational change to include more senior supervision, rotation of junior staff through acute care specialties, a change in the ethos of calling for help, and a formal programme to prepare new doctors for their role in management of the critically ill, including recognition of the early clinical signs of an unstable patient.

Goldhill et al [153] analysed the physiological status of patients immediately prior to unanticipated ICU admission from a hospital ward, and came to similar conclusions. There were deficiencies in recognition of critically ill patients, and deficiencies in acting upon critical physiological derangements. Thirty four per cent of ICU admissions had received cardiopulmonary resuscitation (CPR), and the authors suggested that earlier intervention could have prevented deterioration to this end point. McGloin et al [154] reviewed notes of unexpected inpatient deaths and ICU admissions from hospital wards, and found that patients with obvious clinical signs of deterioration were overlooked or poorly managed, and that this could have contributed to worse outcomes or death after ICU admission. Further studies have reported deficiencies in care in the hours immediately preceding cardiopulmonary arrest [155, 156], and suboptimal inpatient care leading to death or other adverse outcome [10-12, 157].

There is evidence that new graduates may be ill-equipped to deal with the clinical demands of the house surgeon year [158]. The responsibility for teaching critical care does not rest with one discipline, and learning in this domain may be opportunistic and haphazard. Even if formal tuition in lectures and tutorials occurs, the clinical experience that enables students to transfer their theoretical knowledge to practice may be lacking due to the unpredictable occurrence of emergencies and the ethical constraints of allowing students to learn on acutely ill patients. Students must develop a systematic approach to the critically ill, learn to work effectively within a clinical team, and know when to call for help, but opportunities to learn are limited.

This chapter will systematically review the evidence on critical care skills in new graduates, and provide an overview of the current status of teaching in this domain. Two of the author’s studies will be described which add to our knowledge of the use of simulation to improve the ability of new graduates to manage critically ill patients.

4.2 A systematic analysis of critical care skills in new medical graduates

A Medline search was conducted using defined headings to identify articles relevant to adequacy of acute care skills of doctors around the time of graduation, and into the early house officer years. The search results were checked for relevance against set criteria, and the reference lists of all identified articles were searched for additional relevant articles (Figure 1).

Figure 1: Search Strategy

Search terms

Critical Care OR Emergency Care OR Acute Care

AND

Medical Graduates OR Medical Students OR New Graduates OR Junior Doctors

Inclusion criteria	**Exclusion criteria**
Published from 1991 to 2004 inclusive	Whole body simulation used for training or assessment (discussed elsewhere)
Contained original data	Reviews, opinion, conference proceedings, texts, course descriptions.
Primary purpose: competency in critical care	Principle focus: training in ACLS / BLS
Adequacy of critical care education	Focus on early years of undergraduate training or more than two years post registration.
English language	Non-OECD countries (not comparable training)

4.2.1 The evidence on critical care competence of newly graduating doctors

Researchers have used a range of methods to collect evidence from new graduates, including surveys of knowledge of medical management, self-reported experience in acute care skills, and self-reported confidence in procedures.

Smith and Poplett [159] surveyed 77 Senior House Officers (SHOs) and 108 Pre-registration House Officers (PRHOs) on their knowledge of oxygen therapy, ability to interpret pulse oximeter readings, clinical signs of apneoa, signs of impending shock, rates of survival following cardiac arrest, management of the unconscious patient and requirements for obtaining consent (100% response rate). They found significant knowledge deficiencies in all areas, with no difference between the scores of PRHOs and SHOs, suggesting that additional clinical experience during the pre-registration year did not correct the deficiencies. The authors concluded that medical graduates were poorly prepared to identify and treat acutely ill patients.

In a postal survey, Harrison et al [160] asked graduating medical students and new graduates if they were confident they had the knowledge and skills required to manage acute reversible life threatening illness. Of the 101 responders (88% response rate) a high proportion of both groups lacked confidence in their abilities as new house officers. Remes et al [161] asked medical specialists what they thought new graduates should know and be able to do in the domain of acute care, and from this, compiled a list of ten core emergency procedural skills that new graduates should possess. They then asked 504 graduating doctors from five medical schools if they knew the theory of the procedure, had ever performed it, and under what circumstances (80% response rate). Skills included insertion of an IV line, resuscitation of an infant, endotracheal intubation in adult and infant, ventilation by mask, tracheostomy, cardiopulmonary resuscitation, insertion of a chest drain, and pericardiocentesis. Apart from pericardiocentesis, there was a high level of knowledge of the ten procedures, but practical experience was limited to IV access and airway management in adults. The study relied on the accuracy of students' self-reported knowledge and experience, but the authors demonstrated a gap between what specialists expected of new graduates and what the new graduates felt confident to manage.

Other reports have found deficiencies in junior doctors' knowledge of pulse oximetry, fluid balance, and analgesia [162, 163]. Teahon [164] surveyed junior doctors and found that most had not had instruction in drug administration and used incorrect, error-prone techniques. A postal questionnaire of ECG interpretation found that the majority of doctors, including SHOs could not correctly interpret PR and QT intervals, lengthening of which may herald serious disease and imminent arrhythmia [165].

A survey of pre-registration house officers' views on how well their medical school experience had prepared them for their job highlighted the major issue: undergraduate training provided a good knowledge base, but there was a gap between this theoretical preparation and the practicalities of the work required [166].

4.2.2 Adequacy of undergraduate programmes in care of the critically ill

Buchman et al [167] conducted a survey of specialists attending a critical care symposium and found a gap between what student doctors were formally required to learn for their examinations in their institutions, and what the specialists felt the student doctors actually needed to know to effectively intervene in the critically ill patient. Frankel et al [168] conducted a review of medical school curricula in the United States and compared this with the United States Medical Licensing Examination (USMLE) in terms of content on critical care. They found that 45% medical schools included critical care in the curriculum but in 60% this was optional (65% response rate from the 126 schools). Sixty percent of schools taught key critical care procedures and 27% required students to pass an assessment in these skills in order to graduate. In contrast, critical care was prominent in the examination, with 19% of Step II USMLE questions containing critical care components. The implication was that there was considerable diversity in experience in critical care medicine for undergraduates, and the minimal critical care curriculum taught in many medical schools was at odds with the prominence awarded critical care in the examination.

In Europe, Garcia-Barbero and Such [169] surveyed 50 countries regarding critical care teaching in undergraduate curriculum and received replies from 20 countries. The main finding was the diversity of practice in terms of critical care content in the curriculum.

Cook and Smith [170] postulated that there was limited information on examination and clinical assessment of critically ill patients in commonly used medical texts. They reviewed 30 routinely available text books against an agreed set of criteria for assessment of the critically ill patient and found that none had sections devoted to this topic, and few contained a comprehensive, systematic description of assessment of the critically ill patient. Not only do many medical students lack formal instruction and clinical experience in the care of critically ill patients, they also encounter difficulties obtaining this information in written texts.

4.3 Attempts to improve junior doctors' management of critically ill patients

A number of investigators have evaluated the effect of training programmes aimed to improve care of the critically ill. These have included courses of instruction, clinical attachments in intensive care units, emergency medicine departments and rural rotations.

Rogers et al [171] designed a structured programme for 33 fourth year medical students during a four-week ICU attachment, aimed at improving problem solving, decision making and technical skills using a combination of case discussions, simulations and supervised clinical experience. Students sat a written test at the beginning and end of the programme, which was based on clinical scenarios of intensive care patients. There was a significant improvement in mean scores, but as in many other similar studies, a control group was not included, so the contribution of the various components of the programme could not be determined. In addition, there are important differences between the management of the complex ICU patient and the early recognition and management of a gradually deteriorating patient on the ward. DeBehnke et al [172] also showed improved written examination scores in 75 medical students when tested at the beginning and end of an emergency medicine attachment that included case-based learning and clinical experience. The attachment was optional, there was no control intervention, and again, improved scores cannot be ascribed to any particular component of the four-week experience. In both these studies, the outcome measure of written test scores may not predict competence in clinical care.

Culhane et al [173] hypothesised that a rural rotation could improve medical students' confidence in managing life-threatening emergencies, and asked 105 medical students to rate their competency in

12 emergency situations before and after the rotation. (81% response rate). There was an increase of 20% in the number of students reporting themselves as competent in nine of the situations, but many students still did not rate themselves as competent (almost 30% remained unsure of bag valve mask ventilation, and over 65% were unsure how to use a cardiac defibrillator).

A number of other centres have developed formal training programmes in emergency medicine for medical undergraduates and junior doctors [174-177], and many have shown improved examination results following these programmes. Kelly [178] was concerned that this initiative may not necessarily equip students for ward emergencies. He compared scores in a written assessment of two groups of students, 45 of whom had participated in a structured emergency medicine attachment in their fourth year, and were starting their fifth year, and 15 students who were ending their fifth year and had completed fourth year before the introduction of the emergency medicine attachment. The assessment focused on case scenarios that junior hospital doctors could face on the ward and, despite one year less experience, the fourth year students obtained significantly higher scores and committed fewer fatal patient errors than the fifth year group. The students were sampled during a teaching session which was compulsory for fourth year students but not for fifth year students, so the latter group may have been an unrepresentative sample. Despite limitations, this study does address the issue of transfer of knowledge of emergency medicine to care of critically ill ward patients, but again, written tests may not predict competence or performance on the ward.

The “ALERT” (Acute Life-threatening Events – Recognition and Treatment) [179] course is a one-day, multidisciplinary course in the care of the acutely ill patient focusing on a systematic patient assessment, when to call for help, communication, decision making and planning. An evaluation of this course found that junior doctors who had completed the course had improved scores in written assessments [180].

This review of the evidence on critical care education supports the proposition that there is a problem. There is evidence of a deficiency in patient care, evidence that medical students lack knowledge, experience and confidence, and evidence that opportunities for learning how to care for critically ill patients are limited and vary between institutions. Specialist opinion supports this conclusion and the critical care curriculum is at odds with the national examination in the United States. Published attempts to address this gap include four-week attachments to acute care specialties during the undergraduate programme incorporating formal teaching, and these demonstrated improved written test scores over the course of the attachment. There is only limited evidence that these initiatives are relevant to critical care in general hospital wards and no evidence of improved management of the critically ill patient.

There are difficulties in acquiring skills in management of critically ill patients and ensuring junior doctors are competent prior to assuming clinical responsibility. Patient simulation may go some way to overcome these difficulties. Simulation has the potential to supplement haphazard clinical experience, and to provide a standardised, controlled environment for education and assessment in the effective management of the critically ill patient.

4.4 Current place of simulation in the undergraduate medical curriculum in emergency care

A recent world-wide survey identified 158 simulation centres, many of whom were involved in undergraduate education [32]. There are increasing numbers of reports of the use of simulation in the medical undergraduate curriculum.

Patient simulation provides a safe learning environment where events can be scheduled, repeated and observed, offering the potential for greater efficiency and rigour over traditional educational methods. Simulation has in fact been described as an ethical imperative [32, 38, 51]. However, simulation has not yet been widely incorporated into the medical undergraduate curriculum, and exposure of New Zealand undergraduates to simulation centres is extremely limited. One reason for this is the cost of simulation training [181]. Set up costs are high, maintenance and disposables are a major ongoing expense, and adequate staffing with trained instructors and technicians requires considerable resources. In addition, the paucity of evidence supporting the need for simulation centres deters managers from committing to this financial outlay. Does simulation-based teaching warrant the expense?

4.5 Summary of evidence

Student opinion of simulation-based teaching is overwhelmingly enthusiastic [51, 97], but measurable outcomes are difficult to demonstrate. The final measure of effectiveness in medical education should be improved patient outcome, but due to multiple confounding factors, proxy outcome measures may be the best that can be achieved.

Evaluation of the educational process is one such proxy measure, and should be included in all course evaluations. Are the learning outcomes well defined and addressing appropriate needs; are the teaching methods and assessment aligned with the desired learning outcomes; and is the assessment valid and reliable? Simulation has advantages over traditional forms of teaching and assessment in this regard, where alignment of learning outcomes, teaching methods and assessment is easy to demonstrate. A realistic simulation is the closest many students will get to the actual clinical experience, and observing a junior doctor in a simulation may be the closest a supervisor ever gets to seeing them manage a critically ill patient.

However, evidence on effectiveness of patient simulators in undergraduate education is limited and inconclusive, and supporting evidence consists mainly of high ratings of student satisfaction [43, 44, 48, 86]. Morgan et al [55] found scores in performing basic anaesthesia skills improved following a simulator workshop, but detected no advantage over the control intervention of faculty facilitated viewing of an instructional video. Nackman [53] demonstrated improved performance in acute care OSCE stations following simulator workshops, but the control group was historical and received no training.

Much of the published literature on undergraduate simulation used an expensive and resource intensive anaesthesia simulator. A less sophisticated, but much more affordable patient simulator is now widely available, and may be of adequate fidelity to meet the needs at undergraduate level.

4.6 Simulation in undergraduate medical education: bridging the gap between theory and practice [182]

The aim of this study was to evaluate what students learnt from a workshop using a medium fidelity patient simulator, and to seek student opinion on the value of patient simulators in the undergraduate curriculum.

4.6.1 Methods

The subjects in this study were 33 fourth year medical students at the Wellington School of Medicine in New Zealand undertaking the Resuscitation Module in the first half of 2002. The aim of the Resuscitation Module is to learn how to manage medical emergencies. During this four-week attachment students spend time with an anaesthetist in the operating room, attend the emergency department, take part in tutorials and read a basic text on resuscitation. In addition, for the purposes of this study, students attended a three-hour workshop in the National Patient Simulation Centre as part of this module. Maximum class size for the workshop was 11 students.

Ethics committee approval was obtained. Ethical issues were addressed in this workshop at three stages: setting the scene, during the simulation and during feedback after the simulation. Unlike traditional teaching methods, patient simulation encourages students to take an active role in management of an emergency situation. This can be stressful both in terms of participating in a realistically simulated emergency and performing in front of colleagues. Simulator “death” can be particularly stressful and should be avoided, but it would be misleading if inappropriate actions resulted in a good simulator outcome.

To create a safe environment, students were told participation was not compulsory, and assured their performance was confidential and would not count in any way towards their course assessment. Students were reassured that they were not expected to perform beyond their ability as fourth year medical students. Students worked in teams, and chose their own roles within those teams, allowing less confident students to take a minor role in the scenario.

Simulator death was avoided by prompts from the “nurse” who followed directions from the instructor through headphones. The nurse prompted students to initiate treatment when necessary and prevented the students from performing clinically inappropriate actions. For example, the nurse would prompt “Should I give the patient oxygen?” if students had failed to do this within a reasonable time or, “The protocol is to dilute 1 mg of adrenaline in 10 mls and give it one ml at a time in this situation”, if students were about to administer an inappropriate dose. In addition, the instructor observed the scenario through one-way glass and could vary the pace and complexity in response to student performance.

The discussion following the simulations aimed to be collegial and supportive, avoiding blame or humiliation, with a focus on behaviours and not individuals. Positive feedback was emphasised and critique was constructive. Students were encouraged to do most of the talking and identify their own strengths, deficiencies and areas for improvement.

During the workshop, four cases were presented to the students: anaphylaxis, chest pain, coma of unknown cause, and major trauma. The instructor outlined the presenting problem and students were allowed several minutes to plan their response. Students managed their “patient” in groups of five to six, while their colleagues observed next door, through audio-visual relay. A nurse helped the students with tasks such as connecting monitors, and locating medications. During a debriefing, which followed each simulation, students reviewed their performance and clarified diagnostic and treatment pathways.

At the end of the workshop, all students completed an anonymous questionnaire, rating aspects of the session on a scale of 1-5 using questions developed and widely used by the Otago University Higher Education Development Centre. In addition, students scored their level of competency on the material before and after the session on a scale of 1-5 where 1=beginner and 5=master (Appendix 3).

Open-ended questions were included in the questionnaire which sought students' views on the workshop including key learning points, the most valuable aspects of the session and their opinion of simulation in undergraduate medical education.

The medium fidelity Laerdal SimMan patient simulator used for the workshop is a computerised full-body mannikin with a realistic upper airway, chest movement, variable cardiac and breath sounds and a palpable pulse. SimMan can be mask ventilated, intubated, cannulated, given fluids and medications and defibrillated. In the event of a pneumothorax, needle thoracocentesis and chest drain insertion can be performed. Monitors display representations of blood pressure, ECG, oxygen saturation and expired carbon dioxide. The console operator can adjust clinical signs and monitor data as the scenario progresses.

This simulator differs from the more sophisticated and expensive METI or Medsim (Eagle) Human Patient Simulators (HPS) in several ways. The HPS is modelled to respond realistically to changing physiological states or participant interventions such as failure to oxygenate, drug administration or blood loss. In SimMan, the console operator keys in changes in parameters, and there is a limited facility for pre-programmed responses.

Unlike the Laerdal SimMan, the HPS lung model exchanges oxygen, carbon dioxide and anaesthetic agents, which can be measured with standard gas monitors. The HPS can be attached to a ventilator, and will generate realistic airway pressures, tidal volumes and spirometry traces and can be connected to real monitors with appropriate alarm functions. This facility is limited in SimMan. Thus the Laerdal SimMan is a major advance on the relatively unresponsive resuscitation mannikins, but more limited than the HPS in what it can realistically simulate. SimMan is, however, less than 25% of the purchase cost of the HPS, is less expensive to maintain, requires considerably less operator expertise and less programming time is required prior to running clinical scenarios. This makes it a more feasible proposition for purchase and use by medical schools. Also, the additional features of the HPS may be more than is required for many contexts, especially where there is no requirement for measurement of lung volumes and pressures, or interfacing with complex anaesthesia monitors.

In analysing the data generated by the post-workshop questionnaire, descriptive statistics were used for student ratings. The Wilcoxin matched-pairs signed-rank test was used to determine the significance of the change in self-assessed competence rating pre and post workshop. Written responses were coded and grouped into themes by the author.

4.6.2 Results

Students rated the workshop highly and found it a valuable learning experience (Table 11). In particular, they felt comfortable working with the simulator, were encouraged to work as a team and felt the session helped integrate theory with practice.

Table 11: Student evaluation of workshop: Total 33 (100%) responses

Question	*Median*	*IQR*	*R*
Did the instructor create a learning environment in which you felt comfortable?	1	1-2	1-3
How successful was the instructor in encouraging you to work as part of a team?	1	1-2	1-3
Did this session help you to develop confidence to use what you learned in class in the clinical setting?	2	1-2	1-4
Was this session effective in helping you to integrate theory and practice?	1	1-2	1-3
Were the simulations a valuable learning experience?	1	1-2	1-3

IQR=Interquartile Range, R=Range.

Rating scale, 1=very much, very effective or very often. 5=not at all, not at all effective or seldom.

Students felt their competency with the material increased following the workshop (Table 12). The mean increase in competency was 1.11 (SD 0.64), which was statistically significant ($p<0.0001$) using Wilcoxin matched-pairs signed-rank test.

Table 12: Move in competency with material before and after the simulator session

Move in competency	Frequency reported (%)
-0.5	1 (3.13%)
0	2 (6.25%)
1	23 (71.88%)
2	5 (15.63%)
3	1 (3.13%)

1=beginner and 5=master. (32 of 33 participants)

In response to the open questions, 31 students (94%) identified one or more key learning points (Table 13). Thematic analysis of the written comments showed 21 students (64%) felt they learnt more about working in a team, as typified by the comments: “The importance of having a leader”; “don’t be afraid to lead”; the need for “allocating jobs to be done within a team”; “defined roles lead to efficient and safe practice” and the importance of clarity in communication. Students felt they learnt how to implement a more systematic approach to a medical emergency, highlighted by the comments: “The ABC is the best starting point even if the diagnosis seems obvious” and “reinforcement of ABCDE” in the initial assessment of a patient. Students identified a difference between knowing what they should do and actually doing it in a realistically simulated medical emergency, illustrated by the comments “the theory and practice of medicine are two different things”, “[putting] theory into practice is quite difficult”, “[the workshop] reinforced what we’d read” and “provided an opportunity to practice”, and “skills you read about are hard to put into practice in real life.”

Table 13: Key learning points. Students responding (total 31) to each theme

Theme	Students Responding No. (%)	Written comments No. (%)
Behavioural issues	21(64%)	29 (48%)
• Teamwork		15 (25%)
• Leadership		4 (7%)
• Task distribution		5 (8%)
• Communication		3 (5%)
• Effect of stress/managing stress		2 3%
Approach to problem	11(33%)	12 (20%)
• Systematic approach to problem		6 (10%)
▪ Primary response		4 (7%)
▪ Secondary survey		1 (2%)
▪ Need to organise response		1 (2%)
• Need to think broadly re diagnosis		3 5%
• Need to know the basics		3 5%
Learning to put theory into practice	12 (36%)	12 (20%)
Specific knowledge	7(21%)	10 (17%)
• Use of a medication		4 (7%)
• Management of specific problem		3 (5%)
• Use of Oxygen delivery devices		3 5%

No=number of students, % of students responding. Written comments (total 60), number of comments and % of comments falling into each theme and subcategory.

4.6.3 Discussion

In this study we found that undergraduates rated the simulator session very highly. In the workshop, students made all the clinical decisions, planned and administered treatment and arranged investigations and ongoing care. Students' comments confirmed they valued the opportunity to manage a realistic medical emergency on their own and in a safe environment. They could apply their theoretical knowledge of resuscitation to practice, and develop a more systematic response to an emergency without causing patient harm. Students worked on their "patient" in teams, planned their roles in the team prior to the simulation, and reviewed the performance of their own team and those of their colleagues. Because of this, students learnt important teamwork skills, as shown by the high percentage of students who nominated aspects of teamwork as a key area of learning. Despite the stress of performing in front of their peers, students found this to be a comfortable learning environment and wanted more simulator sessions in their medical course. Students felt their level of competency with the material improved following the workshop.

The present study differs from other studies in several ways. Students worked in teams and they managed the clinical problem on their own. The workshop content was aligned with the core objectives of the undergraduate curriculum. Other studies have investigated the acquisition of basic skills in administration of anaesthesia, which is not included in the New Zealand undergraduate curriculum. In this study, students were asked to identify what they learnt from the simulation workshop and to assess their competency before and after the workshop. Information on student experiences in simulator workshops in previous studies has been limited to ratings of items from a questionnaire. Students' written comments in response to open questions provide additional insight into student learning.

Previous studies used a high fidelity human patient simulator capable of simulating the more complex anaesthesia environment. Although the HPS would be of value as a model on which to learn respiratory and cardiovascular physiology, in the present study we showed a similar high degree of satisfaction using a much less expensive, medium fidelity simulator in the context of learning how to manage medical emergencies. Increasing fidelity is related to increasing expense, and the degree of fidelity required to meet the learning outcomes for the particular student group should be considered. The Laerdal SimMan may be adequate for recreating clinical situations outside the operating room or intensive care setting. This makes the use of patient simulators a much more feasible proposition for undergraduate education.

Simulator courses are expensive in terms of faculty time. The workshops in this study required an instructor, a "nurse" and a console operator. On the other hand, the less expensive simulator used in this study is only a fraction of the cost of the high fidelity simulator, is much easier to learn how to operate and may well be as effective for undergraduate training. With lower acquisition costs, many more centres are likely to purchase simulators for undergraduate education. Indeed, more than 22 Laerdal SimMan simulators were distributed in Australasia in 2002 and more than 63 distributed around the United Kingdom (personal communication, Laerdal marketing manager).

Prior to this study, it was unclear what students were learning from simulation workshops. It appears that generic skills such as developing a systematic approach to a problem and learning to work with a team were valued more by the students in this workshop than acquiring items of knowledge or component technical skills specific to the four cases presented. These attributes of systematic problem solving and teamwork are hard to address by other methods and patient simulators may fill a gap in current undergraduate programmes.

This study was limited to student opinion and self-assessment of improved ability to manage medical emergencies. A randomised controlled trial with adequate numbers and appropriate outcome measures could provide objective evidence to support the additional cost of simulation. This study suggests that appropriate outcome measures would include teamwork skills and a systematic approach to the management of medical emergencies. A systematic approach to the critically ill patient became the focus of the next study.

4.7 Patient simulators to improve acute care skills in junior doctors [183]

The literature on the management of critically ill patients suggests there may be deficiencies in training of medical graduates, that physiological data may be ignored, and that a systematic patient assessment may be missing [152-154]. A number of initiatives have been tried to address these deficiencies, including use of patient simulation.

Assumptions of inadequate management skills of new graduates rest on attribution of adverse outcomes to training deficit, evidence of lack of critical care content in the undergraduate curriculum, results of written tests of critical care knowledge and skills and self-reported lack of knowledge, skill or confidence. Simulation may be an appropriate tool to measure competence in critical care in new graduates, and also to measure improvement following training, but evidence is limited.

The studies by Morgan and Cleave-Hogg [55, 99] and Boulet [95] evaluating the psychometric properties of simulation-based assessment, suggest reliable results can be generated when assessing undergraduate skills both in management of problems arising during anaesthesia and in trauma care when using checklists to score performance. In terms of effectiveness, improved anaesthesia skills (in a patient simulator) in medical undergraduates have been demonstrated following training [55] and improved performance in the OSCE stations assessing shock has been demonstrated following training on a patient simulator [53].

The purpose of the following study was to assess the ability of medical undergraduates to initiate management of simulated critically ill patients, to see if this improved following a simulator workshop, and to evaluate their opinion of simulation for training and assessment.

4.7.1 Method

Ethics approval was obtained, information provided to participants and written consent obtained. Over a six month period, 71 medical undergraduates in their fourth and sixth (final) year were scheduled to attend a three hour simulator workshop. We used the Laerdal SimMan full-body computerised mannikin, integrated with monitoring devices, airway and resuscitation equipment.

Students underwent an initial period of familiarisation with the simulator and equipment. In the workshop, students worked in teams in one of three scenarios, each team completing the same scenario twice (baseline and repeat). The scenarios focused on a theme of post-operative shock, looking to develop a systematic approach to assessment of the shocked patient, and reinforce this approach by presenting three patients with shock of different origin. All three scenarios were set in a simulated surgical ward. A faculty nurse was present during the scenarios and provided additional cues on "patient" appearance and assistance with monitoring tasks. Information was available in patient notes and charts when requested. The first five-minute period of each scenario was standardised and scripted to allow consistent scoring of the students' actions and comparison between baseline and repeat scenarios. The scenarios were designed such that immediate action was required within this five-minute time frame. After the initial test period, if required, the faculty nurse could offer suggestions to students to direct them towards appropriate management. This was to ensure that correct treatment was eventually given, the "patient" survived, and the experience was positive for the students.

The baseline scenario (five to seven minutes duration) tested entry level skills. The workshop involved participation in simulated case scenarios, observation of peers, feedback following simulations and a facilitated discussion during which students developed a systematic approach to the shocked patient. The repeat scenarios were of 10-15 minutes duration, with the initial five minutes identical to the baseline. At the end of the workshop students gave written statements in response to a questionnaire seeking their views on learning processes and use of simulation in their assessment (Appendix 4).

The videos of the baseline and repeat scenarios were randomised and assessed independently by two expert examiners not involved with the workshop, who scored the five-minute test period of the scenario. One examiner was a specialist anaesthetist and experienced simulation centre instructor and the second was a senior New Zealand Resuscitation Council instructor. The examiners were blind to the year level of the student team and the order of the scenarios. One instructor had no prior knowledge of the students. An anchored five-point rating scale was used to score three dimensions of performance: systematic approach to the problem, clear leadership and division of tasks between team members. The scores for the three dimensions of performance were averaged, and the mean of the two examiners' scores taken to give a global score for performance in each baseline and repeat scenario. In addition, a checklist score was generated from a list of key clinical management tasks.

Quantitative data from questionnaires were analysed using descriptive statistics and written responses were coded and grouped into themes. The Wilcoxon matched-pairs signed-ranks test was used to compare global scores in the baseline and repeat scenarios and to compare fourth year and sixth year scores. The Mann Whitney U test was used to compare checklist scores of fourth and sixth year students.

4.7.2 Results

All 71 students who attended the workshops agreed to take part in the study. These included 45 fourth year students and 26 sixth year students. All students completed the questionnaire. A total of 21 pairs of baseline and repeat scenarios were scored by the two examiners (Table 13).

On a scale of 1-5, where a score of three equated with adequate performance, the median score in the baseline scenario was 1.83 for the fourth year students and 2.3 for the sixth year students. In the repeat scenarios, median scores improved to 2.67 and 3.33 for the fourth and sixth year students, respectively.

The median scores for the two groups in each of the three scenarios (Table 14), shows evidence of improvement in each of the three scenarios. Combining all pre-test and all post-test scores, overall there was a significant improvement from baseline to repeat ($p<0.001$). Scores were compared to see if there was a difference between the fourth and sixth year students. The median scores for sixth year students were significantly higher than for fourth year students for both baseline ($p<0.01$) and repeat scenarios ($p<0.001$).

Table 14: Scores for fourth and sixth year students for the three baseline and repeat scenarios

		Year of training	**n**	**Median***
Scenario 1	Baseline	4th	5	1.5
		6th	4	2.58
	Repeat	4th	5	2.83
		6th	4	3.58
Scenario 2	Baseline	4th	3	2.0
		6th	3	2.5
	Repeat	4th	3	2.33
		6th	3	3.00
Scenario 3	Baseline	4th	3	1.83
		6th	3	2.83
	Repeat	4th	3	2.33
		6th	3	3.33

n=number of teams rated on a rating scale of 1 to 5 where 1 is poor and 5 is good. *Median score of all the teams at that year level.

The students performed significantly more key tasks in the repeat scenarios than the baseline scenarios ($p<0.05$) (Table 15).

Table 15: Comparison of average checklist scores for baseline and repeat scenarios

	Baseline*	Repeat*	p
Scenario 1	13 (11-18)	17 (14-19)	0.01
Scenario 2	13 (10-16)	16 (11-17)	0.04
Scenario 3	14 (11-18)	15 (13-19)	0.04

(total possible scores 21, 20, 21 for scenarios 1, 2 and 3 respectively)

*Total score (range)

Students compared the process of learning in the simulation workshop with their traditional education in this domain. All 71 students identified advantages, and a number of themes emerged from the written statements (Table 16). Active engagement with the material, and the need to make decisions and to commit to action in a realistic time frame were seen as effective ways of learning and remembering. Students felt they had to “make decisions quickly that really matter” and “look at the findings because you understand the importance.”

Students valued hands-on practice of clinical skills, the use of equipment “to see how things would be done in practice”, and discovered how difficult it was to actually manage a case compared to discussing what should be done, as illustrated by the following student: “Theory’s all very well but you gotta [sic] know how to turn on the oxygen before you can administer it.” The transfer of theoretical knowledge to practice was recognised as an essential step which was facilitated by the simulation workshop: “Even though we may know the theory, it is much different in practice.”

For many students, the simulations provided a memorable and realistic experience from which to learn: “It’s a wonderful way to learn because you remember the situation and the devised plan of attack”; “[The] experience sticks in your mind because you are using all five senses”. Simulation helped students see the relevance of their theoretical knowledge and sort out what was important: with “reading, it is difficult to pick out the key points, here it is obvious.” The simulations identified gaps in students’ knowledge and motivated learning, characterised by this student’s thoughts: “It motivates me to thoroughly learn the theory as I have realised the seriousness of emergencies/resuscitation and the implications of being ignorant.” Seventy out of the 71 students (99%) agreed that they had identified areas that they needed to learn more about, and 67 identified specific areas they intended to address.

Eleven students commented on the stressful nature of the experience. They felt “thrown in at the deep end”, and felt under pressure. However, without exception, these same students linked this feeling to a positive learning experience, describing the value of “hands on experience under pressure”. The stress made the experience more memorable, and in fact, was described as “a good way to get over it”.

Table 16: Thematic analysis of written responses on the process of learning

Themes on learning processes	***No.**
Putting knowledge into practice	16
Active engagement with material	14
Practical nature	13
Learning from experience	5
Opportunity to practice for clinical event	4
Demonstrates relevance, motivates learning	4

*N= number of students responding in each theme.

Most students thought that the simulator was a fair measure of their ability to manage a critically ill patient and 91% thought it should be included in a hypothetical end of year assessment (Table 17).

Sixty-five students gave written responses on the subject of assessment. The realism of the simulation assessment emerged as a prominent theme (n=27). It was “obviously testing what one needs to know,” and students commented that the simulator assessment tested teamwork and the ability to solve problems under pressure. Students recognised limits to fidelity of the simulations (n=12), but also commented that simulation was more valid than written tests. They considered prior training in the simulator to be a prerequisite of any assessment (n=13). Assessment in a team was considered as potentially unfair, but on the other hand, the ability to work in a team was important and reduced the stress of the simulation (n=13). (Numbers in brackets are number of responses in that category).

Table 17: Students’ opinions of simulation as an assessment tool

Simulation-based assessment	**Yes***	**No***	**Unsure***
Good measure of ability	55(80%)	12(17%)	2(3%)
Should be included in assessment	63 (91%)	6(9%)	0

Total 69 responses to each question.

*Number of responses (% of all responses to the question)

4.7.3 Discussion

In this study we found that the ability of final year medical students was assessed as less than satisfactory in the domain of emergency care, and that performance improved following a simulation-based workshop. Students saw simulation as a credible and acceptable assessment tool. The consistently higher scores achieved by more senior students support construct validity of simulation-based assessment in that groups who would be expected to perform better did in fact perform better.

Students found the workshop a valuable method for learning as it was practical, relevant, engaging, helped in the transition from theory to practice and motivated them to learn more about the topic by identifying gaps in their ability. The latter is an important aspect in promoting self-directed learning. Students felt they may be more likely to remember material learnt in this way, which may be related to the engagement, challenge or salience of the experience.

Patient simulators may be useful to assess competence of medical students and previous studies show that reliable scores can be generated. Boulet et al [95] demonstrated reliability and construct validity of simulation-based assessment using highly specified checklists to score acute care skills. Morgan [99] used checklists to score anaesthesia skills in medical undergraduates. Global scores by experts have been shown to be more valid and reliable than checklist scores to assess complex performance, as checklists tend to reward thoroughness rather than competence, and may not allow for alternative approaches to a problem [122]. For assessing overall performance in these highly contextualised scenarios, we considered global scores to be more appropriate than checklists. This study adds to the previous studies by examining global scores of overall performance in the highly relevant context of management of the critically ill patient.

Morgan and Cleave-Hogg [97] found students had a positive attitude to simulation-based assessment of basic anaesthesia skills. We have demonstrated this is also the case in assessment of management of medical emergencies, where 91% of students thought simulation-based assessment should be part of their formal assessment. Students qualified this enthusiasm with concerns around group assessment. On the one hand, students felt assessment in a group may be unfair, but on the other hand, felt it would be too stressful as an individual assessment.

4.7.4 Limitations of this study

It is possible that scores may improve in repeated scenarios through increased familiarity with the simulator and the environment. We attempted to minimise this effect through a period of familiarisation with the simulator prior to the scenarios. Increased student exposure to simulation, or cross-over designs with different test scenarios but similar exposure to simulation could address this problem but were not feasible within the time constraints of the student curriculum.

For a high stakes assessment such as an exit examination, a large number of cases would be required to produce a reliable result for any student [101], and the format used in this study could not reliably rank individual student teams. In addition, the performance of individual students could not be disentangled from that of the team. The combined scores however, provide a meaningful comparison between the two student groups and between baseline and repeat scenarios.

We demonstrated improvement in performance by the end of the workshop, but we are unable to say how long this would persist. Students' comments suggest simulation aids retention of new learning. It would be interesting to know if the acquired knowledge was durable, and if refresher courses were of benefit. This could be a productive area for further study.

4.8 Justification of simulation in medical education

Should medical undergraduates learn how to manage critically ill patients on simulators before they deal with the real event, and should competency be assessed on a simulator? Is there now enough evidence to proceed with creating simulation facilities and modifying the medical undergraduate curriculum and examination? This section will explore the evidence addressing the following questions: is there sufficient theoretical justification and empirical evidence to support simulation; should we seek more evidence; are we looking for the right outcome measures; and finally, is simulation assessment ready for take-off?

4.8.1 Theoretical and empirical evidence to support simulation

The learning processes described by students in the simulation workshops were considered by them to be more effective than traditional educational methods in the context of emergency management. This work sheds some light on why this is the case.

The first study explored what it was that students learnt in simulation workshops in the domain of emergency care [182]. Rather than focusing on specific aspects of medical management, students learnt how to be more systematic in their approach to a problem, how to work together in teams and how to communicate more effectively. These factors have been identified as key elements of management of critically ill patients [179]. This was taken further in the second study, to explore the underlying learning processes in simulation-based education, and student attitudes to simulation-based assessment. Students identified a number of factors that have been associated with effective learning [4, 184, 185].

From the student perspective, simulation can help them to identify gaps in their learning, and motivate them to learn more. These are key factors in promoting self-directed, life-long learning. Kaufman [4] proposed a number of principles to guide teaching practice. These include actively engaging the learner, solving real-life problems, providing opportunities for practice, giving feedback, and facilitating reflection on practice through analysing performance and developing new perspectives and opinions. In simulation-based education, new knowledge is acquired in a clinical context, which may aid subsequent retrieval. The relevance of the learning is immediately obvious to students.

These attributes of simulation-based education are consistent with models of effective educational interventions and widely accepted principles of adult learning. Clearly these principles are applicable to other teaching methods, but the ease with which they can be applied in simulation based education underlies the power of this innovative teaching tool.

From the perspective of the educationalist, simulator workshops enable the direct application of learning theory to teaching practice. The theory of constructivism sees the teacher as the facilitator who provides students with relevant experiences targeted to their level of understanding. Learning occurs where students engage actively with the task provided, and is consolidated by in-depth examination of the new experience. The theory of reflective practice argues that professional competence cannot be achieved through formal teaching, but requires exposure to the “messy” problems of real life. Unexpected events or surprises trigger reflection during the event, so called “learning in action”, while subsequent thinking back on what happened, or “reflection on action” relates the event to prior experiences and consideration of how this may affect future practice. Simulation fits well with these theories.

4.8.2 Should we seek more evidence?

To justify the expense of simulation, it would be desirable to demonstrate that simulation was more effective than alternative teaching methods. Demonstrating outcomes of educational interventions can be problematic. A number of confounding variables influence results, and unlimited access to students for randomised controlled trials is not feasible due to ethical and resource constraints.

The paucity of randomised controlled trials in the medical education literature attests to the problems of conducting them [186]. For example, despite the long history of problem-based learning for medical undergraduates, there is little evidence in support of its effectiveness in improving learning outcomes [187]. It is therefore not surprising that simulation lacks evidence of improved learning outcomes.

Morgan et al [55] compared simulation with video-assisted learning and found no difference in learning effect, where the main outcome measure was the score in an itemised checklist of required actions in the simulator at the end of the workshop. In addition, scores in a written test given 2-30 days after the teaching were compared between groups. However, neither simulation, nor video-assisted learning is standard practice, and in fact there may be little difference in resource requirements. A more useful approach would be to compare simulation with currently used methods. In the second study, we compared simulation with the status quo, or no intervention, a method used in other studies of patient simulation [59]. We considered the baseline performance of final year students nearing completion of their training as a measure of the effectiveness of the traditional approach over their six years of training. A more formal approach using a cross-over, randomised, controlled design using simulation and, for example, a scenario-based tutorial, could potentially provide more convincing evidence.

4.8.3 Are we looking for the right outcome measures?

In evaluating the effectiveness of simulation, the chosen outcome measure needs to be appropriate, and fit for the task. In Morgan's study of medical students' basic knowledge of anaesthesia, the assessment may have been testing little more than the ability to memorise a list of required actions [55]. A number of studies have used written tests [171, 172] or OSCEs [53] to evaluate effectiveness of interventions, which again may not be appropriate in view of the lack of correlation between the results of the different test methods. Observation of real practice would be a more appropriate but problematic method [59], and improved patient outcomes the ultimate measure of an effective intervention in medical education. Perhaps the outcomes we should be measuring from simulator courses should be more aligned with the higher levels of performance such as the ability to apply knowledge in a clinical context, the process of problem solving and the ability to work effectively in a team. The present studies indicate that this is what students learnt from the simulation experience.

Difficulty demonstrating improved outcomes from a PBL medical curriculum may similarly be due to looking for the wrong outcome measures. A test of knowledge in an exit exam or even performance in the first year of graduation may be inappropriate if one considers that the major benefit of PBL may be promoting self-directed learning. This would require tracking students over a longer time period. Long term tracking of outcomes of learning following simulation workshops may also be appropriate, as it has been argued that simulation may assist retention of learning [35].

However, theoretical evidence and evidence collected from students is strongly in favour of simulation. Many current methods of teaching medical students have never been subjected to such scrutiny, or have even been shown to be of limited effectiveness. The benefits as described by large numbers of students are difficult to deny and perhaps it is now time to focus our research on other aspects of simulation learning, rather than design elaborate trials to demonstrate its superiority over other methods.

4.8.4 Is simulator assessment ready for take-off?

A new assessment must be reliable and valid. Boulet [95] has provided some evidence on the psychometric properties of simulation based assessment in acute care skills, though at a fairly basic level, using an OSCE-type format. The current studies have provided additional information on validity, supporting construct validity through superior performance in more senior students, and face validity from the perspective of the students, reflected in their comments that the assessment tested what they needed to know for workplace practice. In addition, it appears that simulator assessment would be acceptable to students, and importantly, that this form of assessment would have a positive effect on learning, in that students identified things they needed to know and felt motivated to learn more about caring for critically ill patients.

However, although it seems likely that a reliable OSCE station could be developed using a simulator, assessment of holistic management of medical emergencies in a team would require further study to determine other aspects of validity, including content validity. Individual versus team assessment could be problematic and the generalisability of the scores would need to be determined to decide on the number of judges and cases that would be required to generate results of sufficient reliability.

4.9 Conclusion

We are now at the point where there is sufficient evidence to incorporate simulation courses in critical care into the curriculum of every medical school. To ensure new graduates can manage a medical emergency on the ward, a simulator assessment that is valid and reliable would be very valuable. Such a tool would also be more appropriate than written tests for evaluating the impact of educational interventions aimed to improve care of the critically ill. The next chapter will describe the testing of such an assessment instrument in the context of anaesthesia.

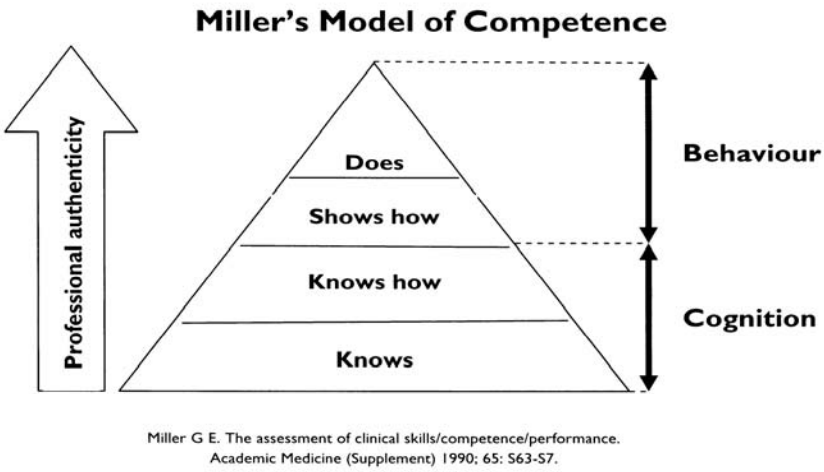
Chapter 5 Psychometric characteristics of simulation-based assessment [188]

5.1 Background

5.1.1 The rationale for performance assessment

The move towards performance based assessment has been in response to highly publicised cases of outstandingly poor performance, and the need to reassure the public that members of the medical profession are competent. There is increasingly closer scrutiny of the processes used in certification, accreditation, and recertification of doctors. Assessments have moved from written tests of knowledge towards methods aimed at measuring what a doctor can do in clinical practice [189]. Miller [190] succinctly illustrated this progression in a diagram often referred to as Miller's triangle (Figure 2).

Figure 2. Miller's Triangle



The Objective Structured Clinical Examinations (OSCE) is widely used to assess competence in a specific clinical task. Standardised patients and simulated patients can be used to assess history taking, physical examination and diagnostic skills. Direct observation of performance in the workplace can be used in the ongoing assessment of trainee doctors and for the revalidation of practicing clinicians. However all methods have their limitations. In an OSCE examination, candidates proceed through a series of stations, performing different tasks. Component parts of performance are measured, but the sum of the parts may not be a valid measure of integrated clinical performance. Where standardised or simulated patients are used in clinical examinations, there can be a problem with the reliability of the scoring, with lack of agreement between different judges, and variable encounters with the patients. Reliability can be improved by standardising the clinical encounter, developing explicit criteria for assessing the performance, and examiner training. However, performance in relation to one clinical problem has been shown to be a poor predictor of performance in a different clinical problem, so that multiple encounters are required. A further limitation with standardised or simulated patients is the restriction of the range of conditions that can be simulated, and the inability to proceed to invasive diagnostic or treatment interventions.

Many training programmes assess trainees by direct observation in the workplace. This should be a valid measure but inadequate training of the examiners, lack of a structured process of observation, and variability of the clinical situation can affect reliability. Ratings have been shown to be unreliable, and suffer from leniency, halo effects (impressions from one item influencing assessment of other items) and restricted use of the grading range [191]. Recent evidence suggests the mini Clinical Evaluation Exercise (mini CEX) may be of value in providing a reliable assessment of trainees in the workplace [192]. A trainee is assessed by a number of different clinical supervisors using a structured rating form, during six or more real clinical encounters spread over a period of six to twelve months. Each assessment lasts 20 minutes and incorporates five minutes of feedback at the end. It is important to ensure that together, the individual assessments in the mini CEX cover the range of required clinical situations.

A variety of approaches has been used in scoring performance, including weighted or unweighted checklists, scales rating overall performance, and time taken to perform key tasks. There is good evidence that global scores of overall performance are more valid and reliable than checklists, where complex clinical ability is being assessed. Regehr [122] demonstrated that when experts rated performance, global rating scales had better reliability characteristics, and were more valid in terms of construct and concurrent validity than checklists. The mini CEX has demonstrated good reliability using global ratings for a number of aspects of performance.

Assessing workplace performance poses particular difficulties in the acute care specialties such as anaesthesia, where an important outcome of training is the ability to manage medical emergencies. The use of simulated and standardised patients is limited to non-procedural aspects of anaesthesia care, and direct observation is limited by difficulty in scheduling critical events, varying degrees of difficulty of the clinical encounter, variable team factors which impinge on the ability of the anaesthetist to manage the patient, and ethical considerations of observing a trainee or colleague without intervening to assist the patient. Patient simulators have attracted attention as a possible assessment tool in anaesthesia. Realistic scenarios can be used to assess aspects of anaesthesia practice that would be difficult to assess by workplace observation, including the ability to effectively manage a crisis. A range of clinical scenarios can be portrayed which can be standardised, repeated, scheduled and videotaped and the observer has no ethical responsibility to intervene for reasons of patient safety.

5.1.2 Evaluating an assessment tool

Any assessment must meet basic requirements to ensure it is fair, fit for the task, and has the desired effect. Assessment can be considered in terms of validity, reliability, feasibility, acceptability, and effect on student learning.

A valid test measures what it intends to measure (face validity), and tests across the range of required knowledge and skills (content validity). External sources of validation should support the results of the test method (concurrent validity), although the “gold standard” measure may be elusive. However, a test that merely replicates the scores of another test adds no value to the assessment process. Groups that one would expect to perform to a higher standard, for example more experienced students or practitioners, should in fact, perform better in the assessment (construct validity). A valid test should predict future workplace performance (predictive validity).

A reliable test should theoretically yield the same result if taken on different occasions and with different examiners. This test-retest reliability is rarely established as it precludes additional learning between tests. Internal consistency measures the correlation between the different items within a test. Errors in scoring can arise due to the variable stringency in ratings by examiners, the so-called “hawks and doves”, or the potential for an interaction between the examiner and the candidate to bias the score, for example gender or racial bias in an examiner. We know that in an OSCE or long case, large numbers of cases and extensive testing time, generally in the order of three hours, are required for an accurate result [101]. Boulet et al assessed acute care skills of medical students and junior doctors using brief scenarios on a patient simulator, and demonstrated that reliable results for examinees could be obtained, but multiple encounters were required [95].

Thus a candidate’s score in an examination will have a number of sources of variance, or error. These are the candidate him or herself, the case, the examiner and the interaction between these components. Ideally we want the candidate to be the largest source of variance, where the purpose of the assessment is to rank the candidates in order of ability. In order to construct a reliable test format, the various sources of error must be quantified.

5.1.3 Assessment of performance in anaesthesia using the patient simulator: the story so far

A good deal of evidence has been collected around the use of simulators for assessment in anaesthesia, which was fully explored in Chapter Two and will be briefly summarised here. Forrest et al [110] reported high levels of inter-rater reliability assessing technical skills of novice anaesthetists against a checklist of required items. Devitt et al and Schwid et al [112, 115] used checklists to assess practising anaesthetists and demonstrated high levels of inter-rater reliability. Gaba et al [120] and Weller et al [121] examined more complex levels of performance, assessing anaesthetists working with colleagues in complex, realistic simulations of anaesthetic emergencies lasting 30-60 minutes, using global rating scales to score performance. Acceptable levels of inter-rater reliability could be obtained only by using multiple raters.

Only one study to date has looked at inter-case reliability in anaesthesia. Murray et al [111] scored the response of junior anaesthetists to a well defined problem in a five minute simulation, and found that large numbers of cases were required to generate reliable results. Devitt et al [114, 116] incorporated sequential, independent problems into a simulation exercise and found low internal consistency of the assessment unless many important items were excluded, suggesting that the performance of the anaesthetists varied between the problems.

Face validity is generally supported by high ratings for realism by participants. A number of studies have shown that up to a certain level, more experienced participants perform better in simulator assessments, supporting construct validity [110-112, 114, 116, 119]. Some studies describe how content validity of their assessment was established [110, 121]. Evidence on concurrent validity of simulator assessment with other tests is limited. Schwid et al [112] found low correlation between simulator assessment scores and scores from traditional assessments.

At this stage, we do not know how many simulator cases should be included in an assessment before we can be confident that the final score truly reflects a candidate's overall ability (in the simulator suite), or the optimum test format, in terms of number of cases and examiners, to produce a reliable, and hopefully feasible, assessment. Little attention has been paid to the feasibility of simulation based assessment, its impact on examinees, its effect on learning, and to what extent it is predictive of workplace performance.

With these large gaps in our knowledge of the properties of simulation-based assessment, there are dangers with premature introduction into certification and recertification processes in anaesthesia. The following study is one part of larger study which aims to address some of these gaps. The remaining parts of the study will be described in Chapter Six.

5.2 Psychometric characteristics of simulation-based assessment of anaesthesia trainees

5.2.1 The aim

The aim of this study was to determine the effect of candidate, case and judge and the interaction between each of these components, on the variance in scores of performance in simulated cases, to determine the reliability of different test formats, and then to estimate the feasibility of using patient simulation to assess performance of anaesthetists in the management of medical emergencies.

5.2.2 Methods

Ethical approval was obtained from the Wellington Ethics Committee. Written informed consent was obtained from all participants and all agreed not to disclose the content of the case scenarios. They were assured that their performance would remain confidential within the study group.

Sampling

There were 22 trainees in the region at the time, with a range of 1-5 years experience in anaesthesia. All had some experience with patient simulation and all were invited to participate in the study.

Scenario development

We used the METI patient simulator placed in a simulated operating room, with an anaesthetic machine and monitoring familiar to the trainees. Three tightly scripted and highly standardised simulator cases were developed, each of 15 minutes duration, involving an emergency developing during the course of an operation. The emergencies were anaphylaxis following induction of anaesthesia, oxygen pipeline failure in a critically ill patient, and cardiac ischaemia followed by cardiac arrest.

The scenarios were developed from those used in a previous study where all three were found to be of a similar level of difficulty [121].

The assessment process

Trainees underwent a structured, 30-minute familiarisation process with the simulator and a scripted introduction to the case. Faculty played the roles of the operating room team and were scripted to give key information at set times and provide help appropriate to their role, and then only in direct response to a request from the trainee. Major events occurred in a set time line, but the course of the scenario could vary appropriately in response to the interventions of the trainee. Prompts from faculty were not permitted. External help could be requested but would not arrive during the course of the scenario.

The trainees took the three scenarios in randomised order to counteract any sequencing effect. There was a break of 10-15 minutes between each scenario. All discussion with faculty was reserved until the study was completed.

Scoring system

We used an anchored, five-point rating scale which had been developed in a previous study by four experienced anaesthetists. In this early study, the draft rating form was applied over the course of ten videotaped test scenarios, and refined through a process of consultation between the anaesthetists to ensure lack of ambiguity and the ability to observe the criteria during the simulations. The form was then used in a study which demonstrated acceptable interrater reliability of performance scores for anaesthetists [121].

In the present study, in addition to the rating form described above, we developed a list of written criteria for expected medical management for each of the three cases, which were agreed on by the study judges. The rating scale had descriptors of each level of performance and in addition, the minimum requirements for an acceptable performance, actions required for an outstanding performance, and actions that would result in a fail were defined.

Assessors gave a global score for overall performance in each of the scenarios based on their expert opinion, guided by the agreed criteria and standards of performance (Appendix 5).

Rating process

To minimise prior knowledge of the New Zealand trainees, we chose judges from Australia. Australia and New Zealand have a joint College of Anaesthetists and identical training requirements, ensuring the judges had similar expectations of trainees. All judges were specialist anaesthetists with experience in simulation. They received a rating package which included an explanation of the study and the rating process, protocols for rating, criteria for assessment, and scoring sheets. Two training meetings and four teleconferences were held as part of rater training. All tapes were rated independently and at no point were judges aware of the scores of other judges for tapes they had not yet rated.

The tapes of the scenarios were de-identified, randomised and rated in the same order at independent sites. All four judges rated the full 15 minutes of all cases.

Statistical Analysis

Results were analysed using generalisability theory with the General Purpose Analysis of Variance (Genova) statistical package. This makes use of all data to quantify all sources of error (subject, case, judge, and interactions between these components), and their relative contribution to the variance in scores [100]. One of the investigators (BJ), an expert in the use of the Genova statistical package, performed this analysis.

The separate sources of variation were combined to express the extent to which differences between trainees reflected reproducible differences. This resulted in a coefficient between 0 and 1 called the generalisability co-efficient (G). A value for G of 0.8 is the accepted level of reliability required for a high stakes assessment.

As the study was a fully crossed, candidate by case by judge design, it allowed generation of the "D" study, or Decision study [100], to determine the psychometric characteristics of different test formats. We could thus determine the most efficient use of cases and judges to produce a meaningful assessment.

5.2.3 Results

Twenty-one trainees participated in the study. One was unable to attend due to service commitments. The tape of one scenario was lost due to recording difficulties, resulting in 62 taped scenarios, which were all rated by the four judges.

The variance components (VC) of scores for trainees, judges and cases and the interactions between them are shown in Table 18. The trainee variance component is an estimate of the variance between trainees' mean scores. As the purpose of the assessment is to detect differences between trainees, most variance should occur here. The case component variance here is zero, indicating there was very low variation in the overall mean scores for cases, and the cases were of equal difficulty. The largest variance component was the interaction between trainee and case, indicating that the rank orderings of different trainees varied with each case. In other words, an individual trainee's performance varied across the three cases. The variance component due to the judge was relatively low, indicating that the judges were of similar average stringency. The variance due to the interaction between judge and trainee was also relatively low, suggesting that the different judges ranked trainees similarly. The final variance component is the total variance in the scores due to the combined effects of judge, case and trainee.

Table 18: Overall performance variance components for candidate, case and judge, and the interaction between these components

Effect under study	Variance component	Standard Error
Trainee	.2758041	.1537267
Case	(0.0)	.0149178
Judge	.0660819	.0660819
Trainee, case	.3712719	.1274148
Trainee, judge	.0575292	.0655307
Case, judge	.0002924	.0195556
Trainee, case, judge (Error)	.7510965	.0986238

The generalisability coefficient for the test format in the study (three cases, four judges) was 0.58, indicating a test format with three cases and four judges is inadequate for a high stakes assessment. With this fully crossed design, (where all trainees took all the cases and all cases were marked by all the judges) G would approach the accepted level of 0.8 only where 10-12 cases were included with three to four judges each rating all the cases (Table 19). From this table it can be seen that increasing the number of cases would increased confidence in the assessment to a much greater extent than increasing the number of judges, reflecting the relatively larger contribution made by variation in trainee performance between cases.

Table 19: Generalisability coefficients (G) for a selection of different numbers of cases and judges

	1 judge	2 judges	3 judges	4 judges	5 judges
1 case	.18	.26	.30	.32	.34
2 cases	.31	.41	.46	.48	.50
3 cases	.39	.50	.55	.58	.6
4 cases	.45	.56	.61	.64	.66
6 cases	.53	.62	.69	.72	.74
8 cases	.58	.69	.74	.77	.78
10 cases	.62	.73	.77	.80	.81
12 cases	.65	.75	.80	.82	.83

Using a nested design for the D study where a different judge marks only a single case, a G of .78 was achieved with 12-15 cases. The implication here is that a test format where a different judge was assigned to mark a single case would be as reliable as four judges marking all cases. Fifteen judges would be required, but they would only have to mark a single case (Table 20).

Table 20: D study for nested design: where one judge is assigned to mark each case

Number of Cases per trainee	Generalisability coefficient
6 cases	.58
8 cases	.65
10 cases	.70
12 cases	.74
15 cases	.77

5.2.4 Discussion

This study used generalisability theory to determine the number of cases and judges required to produce a reliable result for a high stakes assessment. We estimate that three to four hours testing time is required to reliably rank trainees. More is gained by increasing the number of cases than the number of judges, and a single judge marking each case could produce a reliable ranking of trainees if there were 15 cases. This is a more feasible approach than requiring three or more judges to score every trainee in every case.

The results of this study are consistent with findings in other domains where attempts have been made to assess complex aspects of performance. Wass et al [193] carried out a study on final year medical students comparing candidate performance in history taking across two randomly chosen unstandardised cases using a structured rating form. The generalisability coefficient for a single long case with a single judge was only 0.3. They estimated, using generalisability theory, that eight to ten cases would be required to produce generalisability coefficients at or above the acceptable level of 0.8. Similarly, Boulet [95] applied generalisability theory to test the psychometric properties of a simulation-based assessment of acute care skills in medical students and recent graduates and found that six simulations were not enough to produce a reliable result.

This study demonstrates that the case specificity described by Newble [101] applies to assessment of overall performance in simulated anaesthesia crises. The performance of an examinee in one case or clinical simulation is not a good predictor of his or her performance in another case. As in written patient management problems, computer-based simulations, standardised patients and real patient clinical examinations [194-196], the specific knowledge and experience trainee anaesthetists bring to a simulated case seem to be more important than general problem solving skills. Increasing the number of cases is more efficient than having multiple judges rating each case. To increase test efficiency, in the context of OSCEs and written tests, using twice as many cases and one rater per station has been shown to increase reliability to a much greater extent than using two raters per case [101]. This is consistent with the results of this study.

Increasing agreement between judges would reduce the variance component due to the judge. There have been several studies examining this in the context of simulation. It appears that high inter-rater reliability can be achieved where very specific guidelines for evaluation, diagnosis and treatment are specified, but for more complex simulations testing experienced doctors, where there are multiple possible treatment pathways, there is more subjectivity and more variation between judges. In Boulet et al's study [95] assessing the acute care skills of medical students and recent graduates, five-minute cases with specific guidelines for evaluation, diagnosis and treatment were used. High inter-rater reliability of scoring could be achieved objectively as indicated by low variance components for the judge, which suggested that single raters could be used. Morgan and Cleave-Hogg were able to generate reliable assessments of medical students in simulations of anaesthesia problems using checklist scoring of pre-defined actions [99]. In contrast, Gaba et al [120] found that when using complex, lengthy simulations in the context of anaesthesia teams, where performance varied over the course of the simulation, inter-rater reliability was lower. In a previous study, the author found that when assessing experienced practitioners over 30-minute simulations, reasonable reliability could be achieved with three judges [121].

Improving rater agreement is not easily achieved. Gray [197] found that certain rater training methods were effective, such as "frame-of-reference" training, where new raters observe with an experienced rater and discuss their findings. All forms of rater training are likely to require practice and regular feedback to maintain reliable judgements, and may in fact, not be the most efficient and effective way of improving the reliability of an assessment. In this study and others, it appears that multiple judges do not need to mark the same cases, and the focus should be on increasing the number of cases in the assessment.

5.2.5 Limitations and further areas of study

Numbers were limited in this study. It is in fact difficult to generate large numbers as, unlike established assessment methods, there is no existing pool of data, and obtaining data is time-consuming and expensive. It may not prove feasible to recruit external judges willing to rate more than 20 hours of videotaped scenarios. However, including larger numbers of candidates would generate increasingly accurate estimates of G.

In future studies, it may be possible to reduce the length of the scenario, or at least the length of video segment assessed by the judges, to improve feasibility of rating. Intuitively it seems this may result in loss of the ability to assess the more complex aspects of performance, but it may be worthwhile to test this by correlating scores from brief and extended simulations in a cross-over design study.

Numbers of trainees in this study were too small to allow subgroup analysis of performance at different levels of training, or correlate with other markers of performance. It would have been interesting to compare results with those of Devitt et al and Byrne et al [116, 117], who demonstrated a correlation between performance scores and their expectations in terms of experience and training. A study comparing scores of trainees at different levels, or trainees and specialists could provide more evidence on construct validity.

Content validity was not established in this study. For a formal assessment, it would be necessary to develop a bank of cases that tested across the domain of anaesthetic emergencies, establish the individual test item characteristics and ensure each test format was representative.

Concurrent validity has not been assessed in this study. Correlation of simulation scores with examination success would be interesting, but a lack of correlation would not necessarily indicate the simulator assessment was invalid as it may be measuring something different. The correlation of simulator performance and performance in the workplace would be the "gold standard". This may be an unrealistic proposition in the domain of crisis management for reasons previously stated, but it may be possible to compare scores in more routine clinical situations, and there may be components of performance important to both effective crisis management and routine anaesthesia care that could be compared, for example communication, prioritisation and teamwork.

As yet the predictive validity of simulator assessment has not been documented, and it remains to be tested to what extent performance in a simulator will predict future performance in the workplace.

5.3 Conclusion

In this study we have demonstrated that it is possible to reliably rank the performance of anaesthesia trainees using patient simulation. However, a large number of cases is required to produce a result of sufficient reliability for a high stakes assessment.

The above study was designed to test the psychometric properties of simulation used in the summative assessment of trainee anaesthetists. If simulation is to be used in assessment, a number of other aspects require exploration, and in particular, the impact on the learner. The following chapter will explore simulation and assessment from the perspective of the examinee.

Chapter 6 Simulator assessment from the perspective of the candidate, and the accuracy of self-assessed scores [188]

6.1 Introduction

The previous chapter explored the psychometric properties of simulation-based assessment in anaesthesia, and demonstrated that reliable results could be obtained with an appropriate test format. Test reliability is a basic requirement, but a highly reliable assessment is not necessarily a good assessment. Other factors of importance are the validity of the assessment, and the impact it has on student learning. This chapter describes two additional arms to the study of simulation-based assessment of anaesthesia trainees; exploring issues of validity, and the impact on learning from the perspective of the anaesthesia trainee.

6.2 Simulator assessment from the perspective of anaesthesia trainees

Students can offer a valuable perspective on the usefulness of an assessment. Students may see an assessment as a good measure of important attributes, or merely a test of items they see as trivial or irrelevant. There has been a shift in medical education from the traditional teacher-centred approach to a focus on the learner and their needs. It follows that the assessment should also be learner-centred, and the effect of the assessment on the learner investigated. Student opinion of the assessment is an important first step in determining this effect. A student-centred assessment should be seen by students as assessing them fairly, and on items they perceive as important to their future careers. The impact of the assessment on student learning is of major importance, as students are driven to pass examinations, and will focus on learning what is required to pass. Little has been written on simulation-based assessment and its impact on student learning.

6.2.1 Aim

The aim of this arm of the study was to explore simulation-based assessment from the perspective of the anaesthesia trainee, and in particular to answer the question, do trainees find this a reasonable measure of their clinical abilities, and how does it influence their learning?

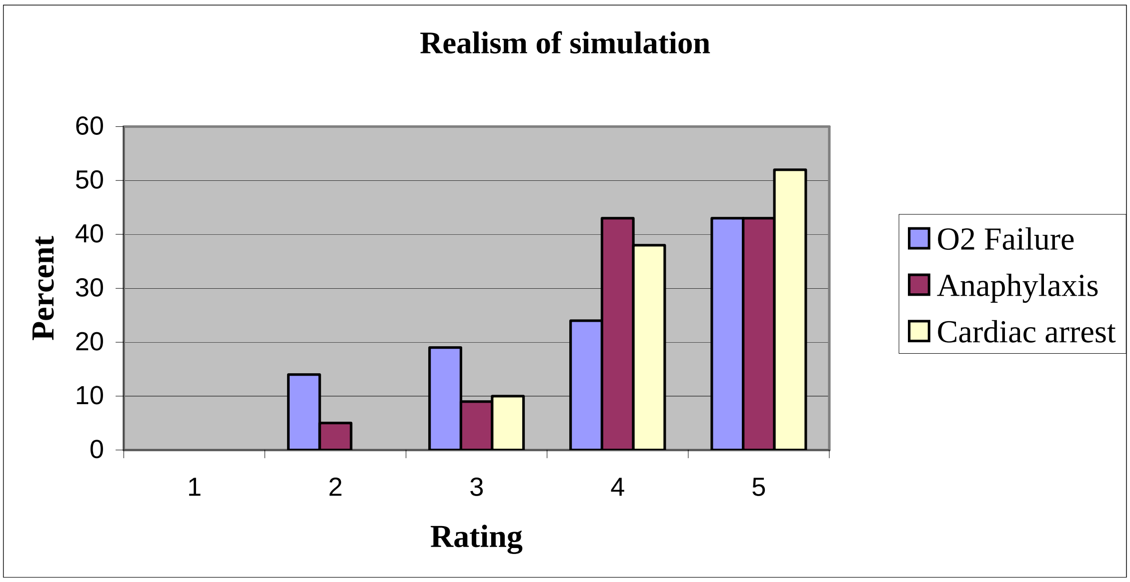
6.2.2 Methods

Twenty-one anaesthesia trainees took part in three simulated cases of anaesthetic emergencies as described in the previous chapter. This group were all anaesthesia trainees in the local district health board (DHB) training scheme at the time and included trainees across all levels of training, with a mix of ten females and eleven males. Following the third simulation, the trainees completed a questionnaire seeking their ratings for the realism of the cases, the factors which improved or decreased realism, the extent to which their performance in the simulator was a reasonable measure of how they would manage the case in real life, and the value of the assessment as a learning experience. The questionnaire was developed by the author for the purpose of the study (Appendix 6). Following completion of the questionnaire, trainees were given the opportunity to discuss the scenario, and receive feedback on their performance. Finally, trainees rated the value of the feedback session. Descriptive statistics were used to analyse questionnaire data.

6.2.3 Results

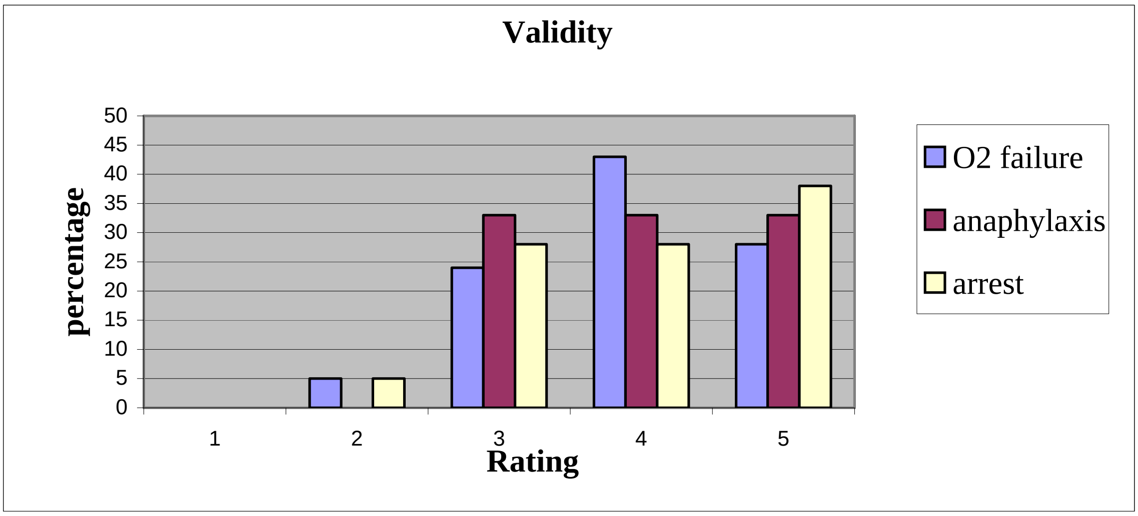
All 21 trainees completed the questionnaire. The realism of the scenarios was perceived as moderate to high by the majority of trainees (Figure 3). All scenarios were rated three or higher for realism by 94% or more of trainees and 64% or more of trainees rated the realism as 4 or 5 on a 5-point scale where five indicated very realistic. Trainees on the whole felt that the scenarios were a good test of their ability to manage these events in real life (Figure 4). The ratings for Oxygen Failure, Anaphylaxis and Cardiac Arrest were three or above in 86%, 95% and 100% of cases, and were rated 4 or above in 67%, 86% and 90% of cases respectively, using a 5-point scale where 5 indicated the simulation assessed their ability very well.

Figure 3: Trainees' ratings of realism of the individual scenarios



1= not realistic and 5 =very realistic.

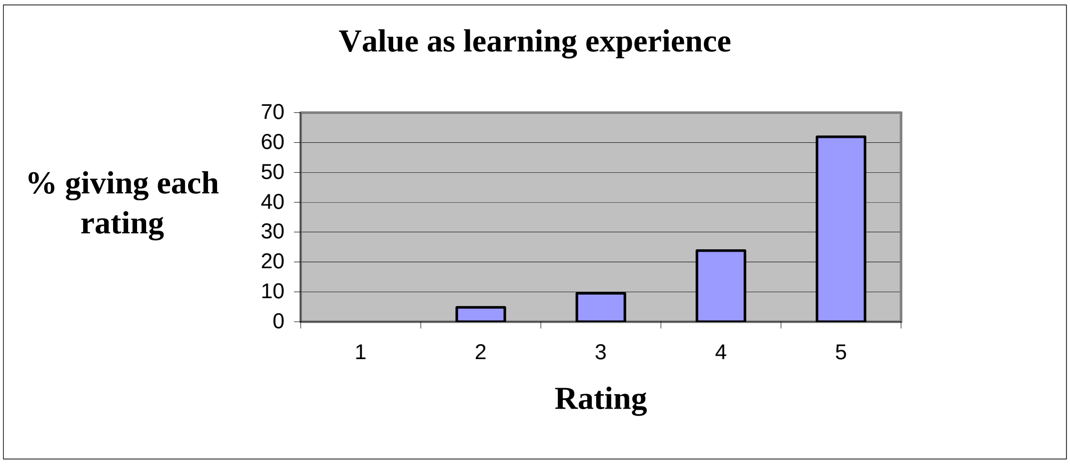
Figure 4: Trainees' ratings of how well the three scenarios assessed their ability to manage the event in real life



1= not at all well and 5 =very well.

Trainees rated the simulator assessment scenarios very highly as a learning experience, where the median rating was 4 and [interquartile range(range)] was [4-5(2-5)] (21 responses, Figure 5).

Figure 5: Trainees' rating of the value of the simulation as a learning experience

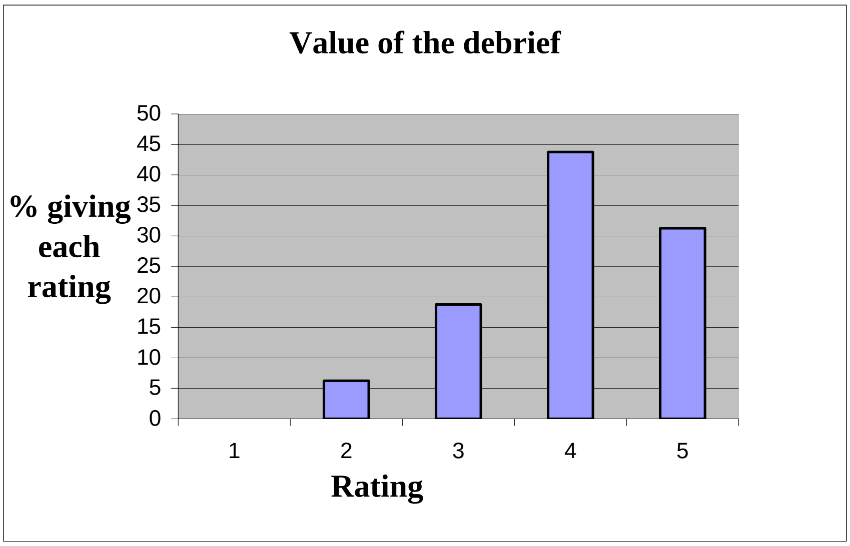


1=not useful, and 5=very useful

Additional written comments on the simulation as a learning experience were offered by ten trainees. These fell into three main themes. Gaining new insight into the difficulties of managing critical events was identified by eight trainees. These related to the experience of working under stress, the number of tasks that needed to be done urgently, how quickly things happened and how rapidly one needed to respond, managing distractions, and trying to keep calm. Two trainees remarked on identifying deficiencies in their performance, and the need to learn more. The second theme, with responses from seven trainees, was the value of the experience as a rehearsal for real life. Trainees valued the opportunity to experience rare events in a realistic and memorable context, and to be forced to manage without direction from their supervisor. The third theme related to teamwork, where three trainees felt they had learnt better how to interact with the team during a crisis.

Responses from 16 trainees were collected on the value of debriefing as a learning experience. Due to opposing service commitments, the debriefing session for five trainees was postponed, and they did not complete the question. Of those 16 who were debriefed at the time, 75% rated the debriefing as 4 or 5 on a scale of 1-5 (Figure 6).

Figure 6. Trainees' rating of the value of the debrief as a learning experience



1=not useful, and 5=very useful

All 21 trainees offered their views on the most and least realistic aspects of the scenarios (Table 21). Participants commented on the realism of the alarms, the feeling of unfamiliarity with the environment, the reality of equipment failures, for example the misassembled self-inflating resuscitation bag, the way the simulator responded to interventions, and the interface with the ventilator. The interaction with the "staff" and the timeline over which events unfolded were found to be realistic. Participants also commented on how their reactions felt realistic, such as "the trap of assuming things", or "losing insight", the "distraction in everyday situation" and "calling for help, but initially thinking you can manage alone".

Comments relating to the least realistic aspects focussed primarily on the features of the simulator mannikin, most notably the difficulty in chest auscultation. In addition, in relation to the oxygen failure scenario, some trainees felt that there were too many difficulties encountered to be realistic. This was in reference to the difficulty in obtaining additional oxygen cylinders in the context of a widespread hospital pipeline failure. Some participants described how their diagnostic approach differed in the simulator, for example, one trainee wrote, “because [I was] in the simulator suite [I] thought, aha, anaphylaxis” and another, “[I was] always expecting the worst to happen, which is not true of real life.”

Table 21: Most and least realistic aspects of the simulations. N=number of responses

Most realistic	**N**
Environment (anaesthetic machine, monitor, drug trolley, airway equipment)	10
Simulator features and response	4
Interaction with actors	4
Scenario script	4
Response to scenario	4
Least realistic	
Simulator features and response	12
Breath sounds	8
Pulse palpation	2
Clinical appearance (airway, skin colour, sweating, perfusion)	3
Monitor data (Oximetry trace, ventilator pressure)	2
Vascular access: ease of insertion of arterial line	1
Acting	3
Lack of back up oxygen cylinders, multiple equipment problems	4
Lack of help	1
Behavioural response (guessing at scenario as simulator, expecting a disaster)	2

6.2.4 Discussion

Trainees found this a realistic experience and a reasonable measure of their ability to manage a similar case in real life. They found the simulator assessment a valuable learning experience, and they identified important items they had learnt.

The perception of trainees that the assessment was a reasonable measure of their ability supports the face validity of the assessment. Their perception of the realism of the simulation also supports face validity, in that it seemed a reasonable representation of real life, demonstrating close alignment between the assessment method and workplace practice. Trainees were asked to consider the elements of the experience that contributed to this sense of realism or otherwise. The emotional responses described suggest a high level of involvement in the simulations. The features of the mannikin and the scenarios which detracted from realism would need to be addressed for a high stakes assessment. The unrealistic approach to problem solving noted by some trainees must also be noted, for example, the expectation that something would go wrong resulting in heightened anticipation, and second guessing the nature of the clinical problem, as a result of previous experiences in the simulator.

Routine exposure to simulation, as in the airline industry, where not all encounters in the simulator develop into a potential crisis, and careful scenario design may go some way to overcoming these problems, but the limitations must be recognised and further explored in order to define what can be meaningfully assessed in the simulator and what requires an alternative approach.

Ideally an assessment should be seen as a credible measure of important attributes, be a positive learning experience and help to identify gaps in learning. As assessment drives student learning, an assessment should encourage anaesthesia trainees to learn how to better manage their patients. Knowledge tests may drive trainees to memorise lists of facts, which may not translate to improved workplace performance. Trainees viewed the simulator assessment as a positive learning experience, they learnt important aspects of crisis management and felt motivated to learn how to better manage these events. This evidence supports the positive impact on learning of simulation-based assessment. There was additional benefit in the process of debriefing after the simulation, which supports the inclusion of feedback into any simulation-based assessment.

There were several limitations in this study. Trainees were aware that their performance in the simulator scenarios would not affect their career progression, and they may not have been as positive in their responses if this had been a high stakes examination, where the additional stress may have limited their ability to learn from the experience. Due to the limited numbers of trainees in this study, it would not have been meaningful to break the group up into subgroups of experience levels to determine the construct validity of this assessment. However, construct validity has been previously demonstrated in similar contexts. No attempt was made to correlate the trainees' performance in the simulator with other measures of their ability, as we demonstrated in the previous arm of the study that to reliably assess the overall ability of an individual trainee to manage an anaesthetic emergency would require more than the three cases used in this study. Furthermore appropriate alternative measures were not readily available. Success in the Part 1 or Part 2 of the Australian and New Zealand College of Anaesthetists Fellowship examination could be one potential measure, but the trainees were all at different stages of their training, some were pre-Part 1 and others were completing their post-fellowship year, and so there was no single examination with which to compare the simulator scores. Correlation with a well-structured and reliable assessment of performance in the clinical environment would be desirable, but no such assessment existed, and creating and administering such an assessment was beyond the scope of this study. These will be considerations for future studies.

This study supports the face validity of simulation-based assessment from the perspective of anaesthesia trainees, who found it a realistic clinical experience and a reasonable measure of their workplace performance. Limitations to realism have been identified, some of which can be overcome by improving the scenario design. The responses from these trainees support the positive impact on learning of simulation-based assessment.

6.3 The ability of anaesthesia trainees to assess their own performance in a simulated anaesthetic emergency

An assessment may be a motivator for further learning by enabling examinees to identify deficiencies in their knowledge or practice. A certain amount of insight on the part of the examinee is required, in order for an accurate self-assessment of personal deficiencies to be made, and this self-assessment is aided to some extent by feedback from others. This ability to make a judgement on one's own performance against external standards of performance is fundamental to continuing professional development of doctors. However, there is considerable evidence that self-assessed scores are inaccurate and correlate poorly with other external sources of assessment, suggesting that doctors are not always able to identify their own learning needs [92, 198].

The extent to which anaesthetists can identify their own deficiencies when tested in clinical scenarios on a patient simulator is largely unexplored. Anaesthetists find the simulated experience valuable and can relate numerous things they learnt, and even describe how their practice subsequently changed. But how much is learnt will depend on the extent to which anaesthetists can recognise their deficiencies, and then take actions to remedy these deficiencies.

6.3.1 Aim

The aim of this arm of the study was to determine the extent to which anaesthesia trainees' self-assessed scores correlated with the assessment by external judges.

6.3.2 Methods

Twenty one anaesthesia trainees managed three simulated anaesthetic emergencies, which were scored by four external judges as previously described. Following the third scenario, and prior to any discussion, the trainees scored their overall performance in each of the three simulations, using the same 5-point rating scale as the judges, with a simple descriptor for each level of performance (1=unsatisfactory, 2=borderline, 3=satisfactory, 4=good, 5=excellent).

The trainees' self-assessed score was compared with the mean of the judges' scores using the Spearman Rank Correlation test. In order to determine the reliability of the mean judges' score, a one-way random-effects ANOVA model was used to determine the correlation of the judges' scores for a single scenario. Chi-squared test was used to compare scores between different groups of trainees.

6.3.3 Results

We obtained both a self-assessed and an externally rated score for all 62 of the recorded simulated cases. One scenario was not taped due to a recording error. Four judges rated 61 of the videotaped scenarios and three judges rated the remaining tape, as the scores of one judge for this single scenario were lost.

An estimate of 0.40, with a 95% Confidence Interval (0.27, 0.54) was obtained for the intra-class correlation for the ratings of the four judges. The estimated reliability of the mean of the four judges' scores for each trainee performance was 0.73, an acceptable level of reliability. There was modest but significant correlation between self-assessed scores and the mean judge score for the 62 scenarios (rho = 0.321, p = 0.01 Spearman rank correlation). In 48 cases (77.4%) the difference between self-assessed and external scores was minimal, with self and external assessment falling within 1 point on the 5-point rating scale. In 15 (24%) of these 48 cases, the scores were the same.

In 12 cases (19.4%) the self-assessed and external scores differed by 1.5-2 points on the scale, and two (3.2%) differed by 2.5 -3 points.

Participants whom the judges scored low overrated their performance, while those the judges scored higher tended to underrate their performance. This relationship is demonstrated in an inverse correlation between the mean judges' score and the difference between the self-assessed and mean judges' score (rho - 0.614, $p < 0.0001$, Spearman rank correlation). In 18 scenarios, the median judges' score was 2.5 or less on the 5 point scale. In 15 (83%) of these low scoring scenarios, the participants awarded themselves a higher score than the judges. In the remaining 44 scenarios, where the mean judges' score was greater than 2.5, only eight (18%) of the participants awarded themselves a higher score than the judges. This distribution was significantly different ($p < 0.0001$, Chi-squared test).

6.3.4 Discussion

We found a modest but significant correlation between trainees' self-assessed scores and the external scores (rho=.321). In high scoring scenarios trainees underrated their performances while in low scoring scenarios trainees overrated their performances.

A number of previous studies have investigated simulation-based assessment in anaesthesia [109, 114-117, 120, 121, 199] and medical undergraduate courses [95, 99, 200], but only one has included self-assessment as a measured item. MacDonald [200], found that medical undergraduates' self-assessed scores of simple technical manipulations were inaccurate, but self-assessment improved with practice of the skill. The current study looks at self-assessment of the more complex performance in doctors in a specialist training programme.

The inaccuracy of self-assessment demonstrated in this study is consistent with evidence from several reviews of the subject [92, 198]. Poor correlation of self-assessed scores with external measures could be expected to improve with increasing knowledge and experience. However, this does not always appear to be the case. Gordon [92] found that self-ratings of global performance over a clinical attachment did not improve with succeeding years of training. While assessment of peers may become more accurate [201], ability to assess oneself remains poor despite advancing knowledge and skills.

The tendency of poor performers to overrate their performances has also been noted in a review by Boud [198], and Gordon [92] found that in the domain of factual knowledge, there was "a vast overconfidence in those who know little". In an attempt to explain this phenomenon, Gordon [92] suggested that self assessment by adult learners is tied to stable self-concepts of general ability, and appears refractory to objective evidence or even judgment by qualified observers.

The ability to accurately assess one's own abilities is important for a number of reasons. Without insight into one's own weaknesses and gaps in learning, there is little incentive to improve. This insight is essential to continual professional development and life-long learning. In a review of the advantages of programmes designed to improve self-assessment, Gordon [202] considered that self-assessment programmes "may promote more mature, collegial and productive learning environments, particularly suited to the training of health professionals".

In the present study, trainees felt they learnt from participation in the scenarios. However, if they could not accurately identify the deficiencies in their performance, some opportunities for learning were missed. The value of a simulated case as a learning experience could be enhanced if self-assessment was more accurate.

Are we overly optimistic in assuming that simulation-course participants will recognise their deficiencies by watching their errors on video or in the group debriefing following the scenario? Interventions such as watching video replay, and even being provided with written performance criteria, have not always been effective at improving self-assessment in other domains in health education [202]. Some interventions do, however, appear to be effective. Martin et al [201] found that showing benchmarking videos of others may improve self-assessment. Boud [198] suggested that communicating to students the specific criteria and standard against which performance is judged and negotiating the criteria for assessment with the students are effective strategies. Explicit discussion and reconciliation of the differences between the self-assessed scores and other sources of assessment data appear to be vital components of programmes to improve self-assessment [202].

The implications for simulation training are clear. It is inefficient and possibly ineffective to rely heavily on doctors to identify their own training needs following participation in a simulated case. Educational benefit will be maximal if the criteria for good and bad performance are negotiated, the expected standards and actions are explicit, and the difference between participants' views of their performances and the view of external observers is reconciled.

In this study, judges were given information on performance criteria and standards, but this was not shared with the trainees, and self-assessment could have been more accurate if trainees were given the additional information prior to scoring themselves [201]. We did not evaluate if accuracy in self-assessment could be improved through post-simulation interventions and there is some evidence that that improved learning outcomes can be achieved by programmes designed to improve skills in self-assessment [201, 202]. Future studies could examine the effect of verbal feedback, video replay or other interventions on trainees' perceptions of their own level of performance.

Differing levels of training and experience with the clinical situation could potentially affect how trainees judged their performance. Previous work in other domains has not demonstrated this to be the case, and in this study numbers were too small to show meaningful differences in trainees at differing levels of experience. This would be an interesting area for future study.

Inter-rater reliability of judges' scores was acceptable at 0.73, but greater reliability would permit more confidence in the findings. It is difficult to generate high levels of inter-rater reliability where experienced doctors are observed over a whole episode of patient care. It could be assumed that checklists are more objective, and could produce more reliable scoring but at this level of performance, it has been demonstrated that global ratings are more valid than checklists, and no less reliable [122].

It is acknowledged that numbers in this study were small, with 62 performance ratings available for comparison. However, this type of data is difficult to generate, as evidenced by the paucity of large studies in the field of simulation assessment. Unlike established examinations such as the OSCE, there is no existing examination process where such data could be accessed.

We have demonstrated a disparity between self-assessed and externally rated scores of performance in the simulator. This raises interesting questions for simulation-based education programmes and future research. Can simulation courses incorporate processes to improve self-assessment? Can such processes help trainees to more accurately assess their performances in the simulator, and can this be transferred to clinical practice?

6.4 Summary

These two studies have increased our knowledge of simulation-based assessment from the perspective of the learner. In contrast to opinions expressed in surveys in the 1990s, from the evidence obtained here it appears that simulation-based assessment may be both acceptable and valid. It also appears that assessment in the simulator environment is likely to have a positive effect on learning, assisting doctors to identify gaps in their practice, as well as providing a tool on which they can practice new skills and behaviours to address these deficiencies.

Feasibility remains problematic, and it is difficult to conceive of a simultaneous simulator assessment of all candidates sitting their fellowship examinations. A more practical, and probably more useful use of simulators may be in regular assessments throughout the training period, and an opportunity for additional practice and reassessment in the case of poor or borderline performance. This could ensure that all trainees sitting for their final examinations had reached the required level of competence in the management of emergencies in the operating theatre.

Chapter 7 Simulation: present and future

7.1 The future utopia of simulation

In the year 2050 medical educators looked back in amusement at the early patient simulators. Technology had advanced to the stage where, to all appearances, all that differentiated the human patient simulators from human patients was the capacity for independent thought. For the purposes of simulating doctor-patient interactions, this deficiency was overcome by sophisticated ventriloquism. As well as the ability to recreate an extensive range of clinical conditions and events, a range of laparoscopic, endoscopic, orthopaedic and minor surgical procedures could be learned on the manikins, which were housed in well equipped simulated teaching hospitals. In many ways, human patient simulation had reached the level of sophistication achieved by aviation simulators at the turn of the century.

Where simulators provided a suitable alternative, it was considered unethical to use patients for the purposes of training new doctors. Medical students and doctors were required to demonstrate competence at the level of simulation prior to being allowed to attempt risky, painful or invasive procedures on a patient. Simulated patient contact began early in the medical programme, and because students gained skills early on simulators, they were more able and more confident to use these skills with real patients when opportunities arose. As a result, all graduating doctors and specialists were competent in the defined core clinical skills, gained first on the simulator and refined on real patients.

The simulated hospital was accessible at all hours, and available for student's self-directed learning. A virtual tutor could guide students through a syllabus of core clinical problems and procedures, taken at the student's pace, and provide feedback on their responses.

Clinical teachers incorporated patient simulation into their lessons across the curriculum, wherever it was perceived as the most appropriate method, and lessons learnt in the simulator were reinforced in the real clinical setting. A teaching ward round would move between the real hospital and the simulated hospital, wherever the opportunities for learning were available.

The availability of simulated patients overcame the variable exposure to the full range of clinical problems, which had been a problem with the traditional apprenticeship model of learning. The ability to schedule, structure and practice events led to increased efficiency, and students learnt faster and more systematically. As a result, expectations of what new graduates were able to do had progressively risen over the previous decades.

The ad hoc nature of the apprenticeship mode of learning had resulted in some doctors graduating or completing specialist training programmes without ever having learnt how to diagnose and manage important clinical problems and in particular, uncommon or life threatening medical emergencies. Numerous studies in the 1990s and the early part of the 21st century had demonstrated increased morbidity and mortality in hospital inpatients as a result of failure to recognise progressive physiological derangement, with inadequate or delayed responses. With the advent of sophisticated simulators, a comprehensive curriculum in management of the critically ill patient had been developed. As a result, the care of these vulnerable patients had improved, with fewer hospital deaths, reduced morbidity and reduced hospital stays.

The need for teamwork had long been recognised as crucial for the effective management of medical emergencies. With the opportunity to train together in the simulated hospital, health professionals became more aware of the roles and perspectives of other team members, improved their communication, rehearsed teamwork skills and effective resource allocation, and became skilled in working together in emergency conditions.

Patient simulators were incorporated into assessment programmes, enabling previously problematic learning outcomes to be tested. Formative assessment was continuous throughout undergraduate and postgraduate training programmes, and accreditation was required for a defined list of procedures and clinical situations before students or junior doctors could begin to treat patients. All doctors were certified as competent to manage a range of procedures and critical medical events relevant to their specialty, and maintained this competence through regular training sessions incorporating patient simulators as part of their recertification requirements. This contributed to improved patient safety, and reinforced public perceptions of an accountable and competent medical workforce engaged in effective self-monitoring and continuous improvement.

As a consequence of the process of ongoing observation and feedback in the simulator linked to programmes of peer review and annual practice appraisal, it was uncommon to hear reports of an underperforming doctor. However, if a doctor's performance caused concern, practice deficiencies could often be diagnosed through observation in the simulated hospital, enabling the development of a targeted remedial programme delivered in the safe environment of the virtual hospital.

Due to the ability to observe medical practice so closely in the simulator, and with the assistance of psychologists, educators, and engineers, the cause of medical errors could be better defined and addressed through education and systems change. A better understanding was gained of how health care teams worked together, providing evidence to guide training in teamwork and communication. Deficiencies in training programmes were identified through common mistakes observed in the simulator and addressed through curriculum change. Interactions between doctors and biomedical equipment could be observed, leading to improved ergonomic design reducing errors in recognition and interpretation of data, and fewer user errors. All new equipment was tested in simulated practice under a range of clinical conditions prior to its introduction into the clinical environment. Interventions designed to improve patient safety could be road tested in the simulator prior to clinical use, for example, innovations in methods of drug delivery, usability of treatment guidelines, and feasibility of emergency response drills.

7.2 Simulated hospital graduates virtual doctors

By the year 2050, a number of external pressures, combined with the availability of highly sophisticated patient simulators, had led to a shift in undergraduate and postgraduate specialty education away from the traditional apprenticeship model. Government directives for increased efficiency and decreased costs of medical treatment meant that it was no longer possible for clinicians to spend time teaching students, or to tolerate the longer times required for non-specialists be involved in patient management. As a result, undergraduates and trainees were discouraged from entering patient care areas. It was considered unethical to use patients for the purposes of training, and patient rights groups demanded treatment by fully qualified doctors. The hospital boards were unwilling to carry the risk of partially trained doctors producing inferior results. In response to the long term shortage of medical practitioners, the length of the medical course had been reduced, as had specialist training times. The simulated hospital was seen by authorities as the solution to all these pressures, and doctors now completed the majority of their training on virtual patients.

A virtual tutor guided students through a syllabus of core clinical problems and procedures at accelerated pace, and incentives were offered for early course completion. There was reduced need for clinical teachers, which was seen as a major advantage by the hospitals, as teaching no longer distracted qualified doctors from their clinical duties.

Medical programme directors now defined their courses entirely in terms of learning outcomes, all of which must be observable and measurable, and students were required to demonstrate competence in all of these defined outcomes in the simulated hospital.

The ability to schedule and standardise patient responses in the simulator led to the development of computer assisted diagnostic programmes, well-defined management protocols, and formalised approaches to team interactions and communication. The shift to protocolised diagnosis and treatment facilitated the development of a new category of health care workers, technicians trained to manage discrete aspects of patient care.

Despite the astounding technical proficiency of new graduates, some of the older clinicians were alarmed at this reductionist approach to medical education, uneasy that important aspects of professionalism were being lost. Some attributes of a medical practitioner were considered too difficult to reliably examine, and so assessments tended to focus on knowledge and skills that could be tested in written tests, OSCEs and simulations. They observed that the ability of graduates to connect meaningfully with their patients was decreasing, and patient-centred decisions in management options were increasingly overruled by prescriptive guidelines. The new doctors were poorly equipped to respond to the messy uncertainties of real medical practice, and responded inadequately to poorly defined clinical signs and novel situations.

7.3 Back to the present

The above scenarios describe the potential contribution simulators may make to medical education and patient safety, through addressing some of the current problems, and also the potential for misguided use of simulation. Simulation is increasingly being used by clinical educators, and it is important that this new teaching method is used wisely, effectively, and based on educational theory and best evidence. The following section summarises the current state of knowledge of simulation, how educational theory can guide its use, and what we still need to know.

7.4 Effectiveness of simulation-based education: current status

The effective use of simulation in teaching will be enhanced if courses and lessons are based on the best evidence in medical education. The use of patient simulators described in these studies aligns well with a number of theories of learning.

Kaufman [4] described a number of principles of effective learning, and participants in the studies identified these principles in relation to the simulation workshops. The learner is actively engaged, solving clinical problems in a realistic time frame. As the clinical scenarios are based on real-life situations, participants recognise the relevance of the material, and are more likely to engage in the lessons. The simulator environment provides opportunities for feedback and practice. Clinical scenarios can be repeated as often as required until the lesson is learnt.

Kolb proposed a theory of experiential learning through a cycle of action and reflection (Kolb's Learning Cycle) [148]. In the studies described here, participants describe both learning from the scenario itself, and from the subsequent debrief. A simulated clinical event represents a memorable experience, on which students can reflect to explore its meaning. The simulation can be videotaped and played back to the learner, facilitating reflection on practice, aided by feedback from the teacher. This consolidates learning through in-depth examination of the new experience. This process of reflection can generate ideas on how lessons learnt in this simulated clinical situation can be generalised to other situations and students can then plan for future experiences. The simulator then provides an opportunity for students to apply and test new learning in subsequent simulated clinical scenarios.

Learning is contextual, i.e. new knowledge acquired in a particular context is more easily recalled in the same context [203, 204]. The realism of the simulations was reported by many study participants. Knowledge learnt in the context of readings or lectures is not easily translated into practice. Simulations of clinical events enable students to learn in a near real-life context, and to apply their theoretical knowledge in a clinically realistic scenario, translating their theoretical knowledge to clinically relevant working knowledge. Simulation has a role in supporting the transition from the pre-clinical to the clinical years, which was recognised by participants in this set of studies.

The theory of constructivism [205] states that new information must be actively processed and built onto existing knowledge frameworks. Learning involves constructing one's own meanings and understanding of reality and the task of the teacher is to provide students with information and experiences targeted to their level of knowledge. The lessons in the simulator can be designed to target the level of the learner, who actively engages with the task. Learning which takes place during the simulation is consolidated by in-depth exploration of this experience in the light of prior knowledge. This active engagement with the task, and the opportunity to design clinical lessons to target learner level is a strength of simulation. The ability to work through a problem without relying on direction from supervisors was described by both medical students and trainees. Having to commit to a diagnosis forces students to process the information they gather and it also identifies weaknesses. Both undergraduates and post graduates noted that they were able to identify deficiencies in their knowledge and practice when put to the test in the simulator. This is an important motivator to further learning, and indeed the specialists reported changes they had made to their practice as a result of attendance at a simulation workshop.

The apprenticeship model of medical education is very effective, as learning is relevant, is centred on patients, occurs in the context in which it will be applied, and incorporates the rich complexity of medical practice. It does however have limitations, some of which can be overcome by simulation. In the safe environment of the simulation centre, mistakes are allowed, and in these studies students appreciated being able to learn in a realistic clinical setting without risk to patients. The unpredictable and ad hoc nature of clinical experience, the inability to schedule events or adjust for complexity to meet the learning needs of students, and time constraints around reflection and feedback in a busy working environment often place limits around learning. This could be addressed by supplementing clinical experience with simulations throughout the medical undergraduate and postgraduate programmes.

In these studies, as in many others reported in the literature, the focus has been on medical emergencies. Undergraduates, postgraduate trainees and specialists recognised the value of patient simulators in learning how to approach the critically ill patient in a systematic way, and in improving the teamwork, communication and leadership required for effective management. Simulation provides an unparalleled opportunity to learn how to manage uncommon or life-threatening events, and improving management of the critically ill patient may be its primary contribution in medical education.

7.5 Simulation and assessment: current status

A good assessment must be reliable. Previous studies have looked at inter-rater reliability, or agreement of judges rating the same scenario. This thesis adds to the body of knowledge on the psychometric properties of simulation in the context of anaesthesia, determining the number of cases, or testing time, required for a stable result, previously described only at the undergraduate level [95]. This thesis demonstrates that a simulation-based assessment of anaesthesia trainees can be a reliable assessment, and defined the optimum test format in terms of cases and judges.

A good assessment must test important attributes, and be acceptable to examinees. Medical undergraduates and postgraduates are keen to learn what they need to know to be good doctors, and the ability to manage an emergency is clearly important. Undergraduates were highly supportive of including simulated cases in their assessment as they felt it tested things relevant to their future work. Anaesthesia trainees found the assessment was a reasonable measure of their real life performance. An assessment seen as relevant is likely to be acceptable to examinees. The positive support for the assessment contrasts with early surveys of anaesthetists, who were of the opinion that simulation was good for training but not for testing. Acceptance by the medical community of this form of assessment in certification, accreditation, and recertification has not been explored in recent years, and it is likely that attitudes have changed where there has been increased exposure to simulation. It may be time to re-evaluate opinion of the medical community.

A good assessment should have a positive impact on learning. Students are, understandably, driven to pass exams and will spend their time on activities that will best achieve this end. Failure to match the assessment to the desired outcomes may lead to distortions in learning. If aspects of the curriculum are not examined, it is likely they will not be learnt. The relative ease of testing factual knowledge encourages a predominance of tests of factual recall. This will encourage superficial, rote learning at the expense of developing abilities in problem solving, synthesis of material or critical evaluation and judgement. If written tests are the principle method of assessment, students are encouraged to spend more time in the library, and less in the clinical environment. Patient simulators may have a positive impact on student learning, as assessments can be more closely aligned with clinical practice, and it is possible to test clinically relevant attributes. This should encourage students to spend more time learning in the clinical environment, acquiring skills in clinical problem solving, decision making and teamwork. In addition, the simulations assisted participants to identify deficiencies in their knowledge and practice. Medical students identified deficiencies in their learning, and realised the significance of these deficiencies and the potential impact on patient care. This is a strong motivator to further learning. Finally, both undergraduates and postgraduates stated that they learnt from the simulator assessment scenarios, a benefit that would be further enhanced by incorporating feedback into the assessment process. If the assessment scenarios themselves are found to be valuable learning experiences this further supports the proposition that simulation-based assessment has a positive impact on learning.

An additional benefit of patient simulators for assessment is the ability to test aspects of the curriculum that would be difficult or impossible to test by other means, in particular, the ability to systematically evaluate and effectively manage a critically ill patient, to lead a team, to prioritise, and to communicate effectively.

7.6 Limitations and gaps

Participants in these studies noted the limitations of simulators, in particular, related to the fidelity of the mannikin. Unrealistic or absent clinical signs detracts from the realism of the simulation experience but increasing fidelity comes at a cost, and the degree of fidelity required will depend on the lesson and the level of the learner. There are few studies evaluating the degree of fidelity required to meet specified learning outcomes, but these studies provide some evidence from participants on what aspects of fidelity detract from the experience. Future development of simulators should be driven by educational needs, defined by participants and educators, and not be left to the manufacturers to decide.

In addition, a further consequence of poor fidelity may be negative learning. Where physical signs are missing on the simulator, for example, sweating and change in skin colour, students may learn to ignore these signs or consider them unimportant. In the simulator, it may be tempting to use shortcuts, for example skipping the anaesthesia machine check, or omitting patient consent and check-in procedures. Bad habits may be encouraged, such as failure to wear protective gloves, re-use of previously opened bags of intravenous fluid, lack of aseptic technique and use of expired drugs.

As noted by participants in these studies, being in the simulator changes diagnostic processes. In a simulator, it can be reasonably assumed that a minor deviation in physiological parameters is likely to progress to a critical event, leading to anticipation and second guessing of what is likely to happen (“Aha! Anaphylaxis”). In addition, clinical reasoning processes may be different in the simulator, and the impact may vary between different groups. Novices formulate diagnoses by a process of data collection and analysis, while an expert may rely more on pattern matching, using a small number of key clinical cues and matching the pattern of presentation against mental models built up from experience. Simulators lack subtlety in terms of clinical signs, and may reward analytic clinical reasoning processes. This may explain why in some studies, the performance of experienced specialists was no better than that of senior trainees. This has implications for using simulation in recertification processes. An alternative hypothesis is that in novel situations, which would include uncommon emergencies, even experienced practitioners revert to analytical diagnostic processes and therefore the test would be fair. This remains to be explored.

A major deficit in our knowledge is the validity of simulator assessment as a measure of real life performance. This will vary with the context, but in the domain of emergency management we would first require a reliable simulator assessment score for a group of doctors. As we now know the required number of judges and cases, this data could be collected. To then reliably score performance of the group in emergency management in the workplace is more problematic and probably not feasible. Alternative sources of validation such as existing tests results could be sought. However, lack of correlation could indicate either that one test was wrong, or that the two tests measured different things. A high correlation would indicate that one of the tests was redundant, i.e. it added no more information about the candidate.

Clearly, simulators are only one of a number assessment tools, and will have a place alongside other methods. Only through multiple sampling methods can an accurate picture of overall ability be obtained. Simulators could add to this picture by testing ability in the domain of emergency management.

7.7 The way forward

Currently, simulation remains in the realm of enthusiasts, often anaesthetists, and is largely unconsidered by the majority of clinical teachers. The next logical step is to define where, across the medical curriculum, simulation could enhance learning and improve patient outcomes.

What is needed is to define the learning outcomes of medical education programmes at undergraduate, trainee and specialist level and against this map, identify where simulation could enhance current teaching methods, or where simulation would be the best method to achieve the desired outcomes. Using simulators in assessment will follow, aligned with the identified learning outcomes and teaching methods.

A curriculum map of learning outcomes, teaching methods and assessment in the various courses and programmes could identify these opportunities. To check that current curricula do in fact address the necessary learning outcomes, an additional approach would be to identify problems or practice deficits in medical care, for example, suboptimal management of unstable inpatients, and where appropriate, use simulation-based courses to fill these gaps.

7.8 Conclusion

There is no incontrovertible proof that simulation is better than other teaching methods, and there may never be proof. Randomised controlled trials are few and far between in medical education, rarely successful, and are generally considered to be an inappropriate research tool for such a complex construct with so many potential confounding factors. Educational research may more usefully be directed to improving learning, rather than proving one method is better than another or undertaking repeated justification studies. There is ample evidence demonstrating that simulators in medical education are effective, more evidence than is available for many other widely used teaching methods.

There is sufficient evidence to support the introduction at both undergraduate and postgraduate level, of simulation-based workshops to improve management of the critically ill patient as a matter of some urgency. Using simulators in formative assessment of acute care skills and medical emergencies appears from the evidence to be justified and could be considered as an imperative, and there is sufficient evidence to use simulators to test specific competencies in acute care skills for purposes of accreditation. For example, the ability to safely induce anaesthesia in a patient at risk of aspiration could be one of a list of competencies required of a new anaesthesia trainee before joining the after-hours roster.

The evidence would support an assessment where students or trainees demonstrated mastery of a range of emergency protocols and drills prior to graduation, for example, anaesthesia trainees should be able to manage cardiac arrest, failed intubation of the trachea, malignant hyperpyrexia, or anaphylaxis. If used to test management of emergencies in an exit examination, sufficient numbers of simulated cases would be required, which may not be feasible. There is some way to go in terms of demonstrating validity of simulators in the area of clinical reasoning, and it would be premature to use simulation to assess diagnostic ability in complex medical problems.

In this thesis I have mapped the field of simulation in medical education and assessment from the level of undergraduate programmes to continuing medical education. The focus of my studies has been on the management of acutely ill patients, in particular in the field of anaesthesia, but many findings can be generalised to other acute care specialities, and indeed to any clinical domain where medical emergencies are encountered.

Appendix 1

Questionnaire on Continuing Medical Education

1. *Grade*: Specialist ☐ MOSS ☐ Other ☐

2. *In what sectors do you work?* (one or more)
Public ☐ Private ☐ Academic ☐ Other ☐

3. *Do you work in a hospital accredited for ANZCA trainees?* Yes ☐ No ☐

4. *How many anaesthetists (excluding trainees) are there in DHB(s) where you work?*
1-3 ☐ 4-10 ☐ 11-20 ☐ More than 20 ☐

5. *How long have you been practising as an anaesthetist ? (excluding training posts)*
0-10 yrs ☐ 10-20 yrs ☐ 20-30 yrs ☐ > 30 yrs ☐

6. *What CME activities have you undertaken during the last 2 years? (one or more of the following)*

Departmental meetings. 1-5 / year ☐ 6-10 /year ☐ more than 10 / year ☐
CECANZ Annual scientific meeting ☐
Single theme meeting. (NZ) ☐
Simulation-based crisis management course. ☐
EMST course ☐
Skills Workshops (airway/TOE, communication skills etc) ☐
ANZCA annual scientific meeting ☐
ASA annual scientific meeting ☐
Other Australian meeting (describe…………………….) ☐
HELP modules. ☐
International conference ☐
Other (please describe) ☐

7. *Conferences include a range of different activities. When attending a conference, what types of sessions do you find most useful?*
…………………………………………………………………………………….
…………………………………………………………………………………….
…………………………………………………………………………………….

8. *In general, how useful do you find the following CME activities in terms of improving your ability to perform your clinical and other professional duties?*

N/A=not applicable

Departmental meetings	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*
CECANZ Annual scientific meeting.	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*
Single theme meeting. (NZ)	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*

Simulation-based Crisis Management course.	*Very useful*	1	2	3	4	5	*not at all useful*	*N/A*
EMST course.	*Very useful*	1	2	3	4	5	*not at all useful*	*N/A*
Skills Workshops	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*
Australian conferences.	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*
International conferences.	*Very useful*	1	2	3	4	5	*not at all useful*	*N/A*
HELP Module	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*
Other (please identify)	*very useful*	1	2	3	4	5	*not at all useful*	*N/A*

Comments:
……………………………………………………………………………………….
……………………………………………………………………………………….

9. One purpose of CME is to improve patient outcomes by changing physician practice and improving skills. *What changes have you made to your practice or skills as a result of your involvement in a CME activity in the last 2 years?*
……………………………………………………………………………………….
……………………………………………………………………………………….

10. *What CME activity was responsible for the change you describe in question 9?*
……………………………………………………………………………………….
……………………………………………………………………………………….

11. *Do you receive the HELP modules?* Yes ☐ No ☐

If yes: a) *Do you read them?* Always 1 2 3 4 5 never

b) *Do you submit answers for MOPS points?*
Always 1 2 3 4 5 never

c) *How useful do you find them?*
Very 1 2 3 4 5 useless

12. Educational publications come in many forms including clinical updates, systematic reviews, practice guidelines, (often with quiz) electronic journal club, HELP modules and many others either web or paper based.

a) What type educational publication would you prefer?
……………………………………………………………………………………….

b) Would you like to submit answers for MOPS points? yes ☐ no ☐

c) *Would you prefer the format to be*: electronic ☐ written ☐ both ☐ don't care.☐

13. *What motivates you to participate in CME activities?*
……………………………………………………………………………………….

14. *Are you satisfied that your current CME activities are adequate for your clinical and professional development?*

yes ☐ No ☐

15. *What are the* ***major*** *impediments to your participation in CME?*

a) Too many other work commitments. ☐

b) Distance to meetings. ☐

c) Difficulty getting time off. ☐

d) Lack of funding. ☐

e) Lack of relevant CME activities. ☐

f) Other. ☐

Please describe.

…………………………………………………………………………………………….

16. *What specific areas of anaesthesia practice would like to see included in future CME programmes.*

…………………………………………………………………………………………….
…………………………………………………………………………………………….

17. Comments or suggestions.

…………………………………………………………………………………………….
…………………………………………………………………………………………….

Thank you for your time.

Appendix 2

Post EMAC Course Evaluation

Dear colleague,

We would appreciate your response to the following questionnaire, which is intended to evaluate the effectiveness of the EMAC course. Your response may be emailed or s posted (the latter will retain anonymity)

Please answer the following questions:

1. How relevant did you find the course to your clinical practice?

Not relevant				*Very relevant*
1	*2*	*3*	*4*	*5*

2. What were the key points you learnt from the course?

3. What have you found of most value to your clinical work?

4. To what extent did the course affect your confidence in managing crises in the theatre?

Decreased	*No change*	*Maybe Increased*	*Increased*	*Increased a lot*
1	*2*	*3*	*4*	*5*

5. Subsequent to the course, describe any changes you have made to your routine practice, communication or interactions with operating theatre team. (if any)

6. What were the key points of crisis management you learnt in the course?

7. Have you implemented any of these in your clinical practice? *Yes* *No* *Unsure*

If yes, please describe.

8. Have you been involved in managing a crisis in theatre subsequent to the course? *Yes* *No*

If yes, Please describe;

9. Was your management affected in any way as a result of participating in the EMAC course? *Yes* *No* *Unsure*

If yes, please describe:

8. Do you plan to attend further EMAC courses in the future? *Yes* *No* *Unsure*

If yes, what would be your main purpose for returning? (apart from the great food, company, setting)

Comments

Thank you for your time

Appendix 3

Questionnaire
Crisis Management Workshop

Date:
Regarding the session on crisis management:

1. How organised did you find this session?	Well organised	1	2	3	4	5	Poorly organised
2. How would you rate the instructor's ability to communicate ideas and information?	Excellent	1	2	3	4	5	Poor
3. How much did this session stimulate your interest in the subject?	Very much	1	2	3	4	5	Not at all
4. How would you describe the instructor's attitude to students in this session?	Very helpful	1	2	3	4	5	Not at all helpful
5. Overall, how effective was the instructor in teaching this session?	Very effective	1	2	3	4	5	Not at all effective
6. Did the instructor create a learning environment in which you felt comfortable?	Very much so	1	2	3	4	5	Not at all
7. How successful was the instructor in encouraging you to work as part of a team?	Very successful	1	2	3	4	5	Not at all successful
8 Did this session help you to develop confidence to use what you learned in class in the clinical setting?	Very much so	1	2	3	4	5	Not at all
9. Were you encouraged to think through clinical problems for yourself?	Very often	1	2	3	4	5	Seldom
10. Was this session effective in helping you to integrate theory and practice?	Very effective	1	2	3	4	5	Not at all effective
11. I found the simulations a valuable learning experience.	Strongly agree	1	2	3	4	5	Strongly disagree

12. Rate your level of competency on the material covered in this session:

Before the session	Beginner	1	2	3	4	5	Master
Now	Beginner	1	2	3	4	5	Master

11. What were the key things you learnt from this session?

12. This session consisted of case planning, simulation, discussions, presentations and repeating the simulation. What aspects of this session did you find most valuable in terms of learning this material?

13. What did you like least about this session?

14. What is your opinion on the use of the simulator in undergraduate medical education?

Thank you for your feedback.

Appendix 4

Student Questionnaire
A Simulation Based Workshop on the Management of Shock

1. What year are you in at medical school?
   Year 1, 2, 3, 4, 5, Trainee Intern. (*circle one*)

2. Was this workshop pitched at the right level for you? Yes ☐ Too Basic ☐Too complex ☐
   Comments:

3. At what year of medical school do you think you would gain most benefit from this workshop?
   Year: 1, 2, 3, 4, 5, Trainee Intern. (*circle one*)

4. What are the key things you have you learnt from this workshop?

5. What is the main thing that sticks in your mind from this session?

6. How could this workshop have been improved?

7. As a way of learning, how does a simulation workshop on management of shock compare to case-based learning, lectures or reading of the same topic? What do you see as the advantages, disadvantages or differences?

8. Have you identified any aspects of management of the shocked patient where you feel you need more knowledge or skills? Yes ☐ No ☐If yes, what are they?

9. How do you intend to address the issues you identified in Question 8 (if any)?

10. How do you feel about your own performance in the scenarios?

11. At a hypothetical end of year assessment, do you think the simulator could provide a fair measure of your ability to manage a critically ill patient? Yes ☐ No ☐

    Explain your reasons for or against.

12. Do you think the simulator should be included in assessment at some stage during medical school?
    Yes ☐ No ☐ If yes, at what year level(s)

    Yr: 1 2 3 4 5 TI

Thank you for your time.

Appendix 5

Global Rating Form
Simulator Generalisability study

Rater:………………………………………

Date:…………………Time: ……………..

Tape Code:……………………………….Participant gender: M ☐ F ☐+

Scenario: Anaphylaxis ☐ O2 Failure ☐ Arrest ☐

Refer to scenario specific checklist for medical management.

1=fail, 2=borderline, 3=acceptable, 4=good, 5=outstanding

Medical Management	1	2	3	4	5
Treatment is timely.					
Gathers the appropriate information **.**					
Reaches diagnosis, considers differential.					
Institutes appropriate treatment .					
Appropriate order of interventions.					

Behaviour	1	2	3	4	5
Acquisition of all available information					
Anticipates and plans					
Calls for help appropriately					
Re-evaluates situation					
Utilises team resources effectively					

Allocates attention wisely

Prioritises

Concise, directed instructions, closes communication loop

Communicates problem clearly to team, listens to team

Overall Peformance	1	2	3	4	5

Key issues with performance:

(best and worst aspects)

***Score justification**:*

***Technical problems with videotape**:*

***Other comments**:*

Appendix 6

Post-Study Participant Questionnaire

Thank you for your participation. Could you please answer the following questions.

1. How realistic did you find the three scenarios?

(where 1=not realistic and 5 = very realistic)

O2 failure	1	2	3	4	5
Anaphylaxis	1	2	3	4	5
Cardiac arrest.	1	2	3	4	5

2. How well do you think the scenario assessed your ability to manage these events in real life?

(Where 1=not at all well, 5 = very well)

O2 failure	1	2	3	4	5
Anaphylaxis	1	2	3	4	5
Cardiac arrest.	1	2	3	4	5

3. How would you rate your performance in the three scenarios?

Where 1=unsatisfactory, 2=borderline, 3=satisfactory, 4=good, 5=excellent

O2 failure	1	2	3	4	5
Anaphylaxis	1	2	3	4	5
Cardiac arrest.	1	2	3	4	5

4. How would you rate your participation in the simulator scenarios (not including the debrief) as a learning experience?

Not useful 1 2 3 4 5 very useful

5. What were the elements of this learning experience (the scenario) that had the most impact on you?

6. How useful was the debriefing as a learning experience (where relevant)

Not useful 1 2 3 4 5 very useful

7. What were the elements of this learning experience (the debriefing) that had the most impact on you?

8. What aspects of the scenarios did you find most realistic and least realistic?

References

1. GMC, *Tomorrows doctors*. 2003. p. www.gmc-uk.org/med_ed/tomdoc.htm.

2. WFME, *World Federation for Medical Education*, http://www.sund.ku.dk/wfme/top.htm, Editor.

3. Swing, S., *Assessing the ACGME general competencies: general considerations and assessment methods.* Academic Emergency Medicine, 2002. **9**(11).

4. Kaufman, D., *Applying educational theory in practice.* British Medical Journal, 2003. **326**: p. 213-216.

5. Cooper, H., C. Carlisle, T. Gibbs, and C. Watkins, *Developing an evidence base for interdisciplinary learning: a systematic review.* Journal of Advanced Nursing, 2001. **35**(2): p. 228-37.

6. HWAC, *Health Workforce Advisory Committee: Fit for purpose and fit for practice: a review of the medical workforce in New Zealand consultation document*. 2005.

7. Ericsson, K. and N. Charness, *Expert Performance: Its Structure and Acquisition.* American Psychologist, 1994. **49**(8): p. 725–747.

8. Guest, C., G. Regehr, and R. Tiberius, *The life long challenge of expertise.* Medical Education, 2001. **35**: p. 78-81.

9. Eva, K., *The aging physician: changes in cognitive processing and their impact on medical practice.* Academic Medicine, 2002. **77**(10 Supplement): p. S1-S6.

10. Vincent, C., G. Neale, and M. Woloshynowych, *Adverse events in British hospitals: preliminary retrospective record review.* British Medical Journal, 2001. **322**(7285): p. 517-519.

11. Leape, L., T. Brennan, N. Laird, A. Lawthers, A. Localio, B. Barnes, L. Hebert, J. Newhouse, P. Weiler, and H. Hiatt, *The nature of adverse events in hospitalized patients: The results of the Harvard Medical Practice Study II.* New England Journal of Medicine, 1991. **324**: p. 377-384.

12. Wilson, R., W. Runciman, R. Gibberd, B. Harrison, L. Newby, and J. Hamilton, *The Quality in Australian Health Care Study.* Medical Journal of Australia, 1995. **163**: p. 458-471.

13. Davis, P., R. Lay-Yee, S. Schug, R. Briant, A. Scott, S. Johnson, and W. Bingley, *Adverse events regional feasibility study: indicative findings.* New Zealand Medical Journal, 2001. **11**(114): p. 203-205.

14. Bognor, M., *Human Error In Medicine*. 1st ed. 1994, New Jersey: Lawrence Erlbaum Association Inc. 411.

15. Helmreich, R. *Threat and error in aviation and medicine: Similar and different*. in *Special Medical Seminar, Lessons for Health Care: Applied Human Factors Research*. 2000. NSW: Australian Council of Safety and Quality in Health Care & NSW Ministerial Council for Quality in Health Care.

16. Merry, A. and A. McCall Smith, *Errors, Medicine and the Law*. 2001: Cambridge University Press.

17. Reason, J., *Human Error*. 1st ed. 1990, Cambridge: Cambridge University Press.

18. Endsley, M., *Toward a theory of situation awareness.* Human Factors, 1995. **37**: p. 65-84.

19. Fletcher, G., R. Flin, P. McGeorge, R. Glavin, N. Maran, and R. Patey, *Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system.* British Journal of Anaesthesia, 2003. **90**(5): p. 580-588.

20. Gaba, D., K. Fish, and S. Howard, *Crisis Management in Anesthesiology*. 1st ed. 1994: Churchill Livingston. 290.

21. Runciman, W., A. Sellen, R. Webb, J. Williamson, J. Currie, C. Morgan, and W. Russell, *Errors, incidents and accidents in anaesthetic practice.* Anaesthesia and Intensive Care, 1993. **21**(5): p. 684-694.

22. Reznek, M., H. P, and T. Krummel, *Virtual reality and simulation: training the future emergency physician.* Academic Emergency Medicine, 2002. **9**(1): p. 78-87.

23. Morag, R.M., S. DeSouza, P.A. Steen, A. Salem, M. Harris, O. Ohnstad, J.T. Fosen, and B.E. Brenner, *Performing procedures on the newly deceased for teaching purposes: what if we were to ask?* Archives of Internal Medicine, 2005. **165**(1): p. 92-6.

24. Hansen, L.A. and G.R. Boss, *Use of live animals in the curricula of U.S. medical schools: survey results from 2001.* Academic Medicine, 2002. **77**(11): p. 1147-9.

25. Willis, L.R. and H.R. Besch, Jr., *Effect of experience on medical students' attitudes toward animal laboratories in pharmacology education.* Academic Medicine, 1995. **70**(1): p. 67-9.

26. Denson, J. and S. Abrahamson, *A computer controlled patient simulator.* JAMA, 1969. **208**: p. 504-508.

27. Gaba, D., *Improving anesthesiologists' performance by simulating reality.* Anesthesiology, 1992. **76**(4): p. 491-494.

28. Gaba, D. and A. DeAnda, *The response of anesthesia trainees to simulated critical incidents.* Anesthesia and Analgesia, 1989. **68**: p. 444-451.

29. Abrahamson, S., J. Denson, and R. Wolf, *Effectiveness of a simulator in training anesthesiology residents.* Journal of Medical Education, 1969. **44**: p. 515-519.

30. Issenberg, S., W. McGaghie, and E. Petrusa, *Features and uses of high fidelity medical simulations that lead to effective learning: a BEME systematic review.* Medical Teacher, 2005. **27**(1): p. 10-28.

31. Kapur, P. and R. Steadman, *Patient simulator competency testing: ready for takeoff?* Anesthesia and Analgesia, 1998. **86**: p. 1157-1158.

32. Morgan, P. and D. Cleave-Hogg, *A worldwide survey of the use of simulation in anesthesia.* Canadian Journal of Anaesthesia, 2002. **49**(7): p. 659-662.

33. Chopra, V., F. Engbers, M. Geerts, W. Filet, J. Bovill, and J. Spierdijk, *The Leiden anaesthesia simulator.* British Journal of Anaesthesia, 1994. **73**: p. 287-292.

34. Arne, R., F. Stale, K. Ragna, and L. Petter, *PatSim- simulator for practising anaesthesia and intensive care.* International Journal of Clinical Monitoring and Computing, 1995. **13**: p. 147-52.

35. Issenberg, M., W. McGaghie, I. Hart , J. Mayer, J. Felner, E. Petrusa, R. Waugh, D. Brown, R. Safford, I. Gessner, D. Gordon, and G. Ewy, *Simulation technology for health care professional skills training and assessment.* JAMA, 1999. **282**(9): p. 861-866.

36. Friedrich, M., *Risk-free medical training with patient simulators.* JAMA, 2002. **288**(22): p. 2808-12.

37. Schwid, H., *Anesthesia simulators-technology and applications.* Anesthesiology, 2000. **2**: p. 949-53.

38. Ziv, A., P. Wolpe, S. Small, and S. Glick, *Simulation-based medical education: an ethical imperative.* Academic Medicine, 2003. **78**(8): p. 783-788.

39. Maran, N. and R. Glavin, *Low- to high- fidelity simulation - a continuum of medical education?* Medical Education, 2003. **37 (suppl 1)**: p. 22-28.

40. Gordon, J., N. Oriol, and J. Cooper, *Bringing good teaching cases "to life": a simulator-based medical education service.* Academic Medicine, 2004. **79**(1): p. 23-7.

41. Wong, A., *Full scale computer simulators in anesthesia training and evaluation.* Canadian Journal of Anaesthesia, 2004. **51**(5): p. 455-64.

42. Larbuisson, R., P. Pendeville, A. Nyssen, M. Janssens, and A. Mayne, *Use of anaesthesia simulator: initial impressions of its use in two Belgian University centers.* Acta Anaesthesiologica. Belgica, 1999. **50**(2): p. 87-93.

43. Good, M.L., *Patient simulation for training basic and advanced clinical skills.* Medical Education, 2003. **37 (Suppl. 1)**: p. 14-21.

44. Tan, G., L. Suresh, B. Ho, and T. Lee, *Teaching first-year medical students physiology: does the human patient simulator allow more effective teaching?* Singapore Medical Journal, 2002. **43**(5): p. 238-242.

45. Euliano, T., *Teaching respiratory physiology: clinical correlation with a human patient simulator.* Journal of Clinical Monitoring and Computing., 2000. **16**(5-6): p. 465-70.

46. Euliano, T. and M. Mahla, *Problem-based learning in residency education: a novel implementation using a simulator.* Journal of Clinical Monitoring and Computing, 1999. **15**(3-4): p. 227-32.

47. Euliano, T., *Small group teaching: clinical correlation with a human patient simulator.* Advances in Physiology Education, 2001. **25**(1): p. 36-43.

48. Darin, K., R. Kyle, J. Trask, C. Shields, and P. Mongan, *Using high-fidelity patient simulation and an advanced distance education network to teach pharmacology to second-year medical students.* Journal of Clinical Anesthesia, 2004. **16**: p. 144-151.

49. McIvor, W., *Experience with medical student simulation education.* Critical Care Medicine, 2004. **32**(2): p. S66-S69.

50. Grenvik, A., J. Schaefer, M. DeVita, and P. Rogers, *New aspects on critical care medicine training.* Current Opinion in Critical Care, 2004. **10**: p. 233-7.

51. Gordon, J., W. Wilkerson, D. Shaffer, and E. Armstrong, *"Practicing" medicine without risk: students' and educators' responses to high-fidelity patient simulation.* Academic Medicine, 2001. **76**(5): p. 469-472.

52. Flanagan, B., D. Nestel, and M. Joseph, *Making patient safety the focus: Crisis resource management in the undergraduate curriculum.* Medical Education, 2004. **38**: p. 56-66.

53. Nackman, G., M. Bermann, and J. Hammond, *Effective use of human simulators in surgical education.* Journal of Surgical Research, 2003. **115**: p. 214-8.

54. Cleave-Hogg, D. and P. Morgan, *Experiential learning in an anaesthesia simulation centre: analysis of students' comments.* Medical Teacher, 2002. **24**(1): p. 23-26.

55. Morgan, P., D. Cleave-Hogg, J. McIlroy, and J. Devitt, *Simulation technology: a comparison of experiential and visual learning for undergraduate medical students.* Anesthesiology, 2002. **96**(1): p. 10-16.

56. Hammond, J., M. Bermann, B. Chen, and L. Kushins, *Incorporation of a computerized human patient simulator in critical care training: a preliminary report.* Journal of Trauma, 2002. **53**(6): p. 1064-7.

57. Lee, S., M. Pardo, D. Gaba, Y. Sowb, R. Dicker, E. Straus, L. Khaw, D. Morabito, T. Krummel, and M. Knudsen, *Trauma assessment training with a patient simulator: a prospective, randomized study.* The Journal of Trauma, Infection and Critical Care, 2003. **55**(4): p. 651-657.

58. Marshall, R., S. Smith, P. Gorman, T. Krummel, R. Haluck, and R. Cooney, *Use of human patient simulator in the development of resident trauma management skills.* The Journal of Trauma, Infection and Critical Care, 2000. **51**(1): p. 17-21.

59. Mayo, P., J. Hackney, J. Mueck, V. Ribaudo, and S. RF, *Achieving house staff competence in emergency airway management: results of a teaching program using a computerized patient simulator.* Critical Care Medicine, 2004. **32**(12): p. 2422-7.

60. Olympio, M., R. Whelan, R. Ford, and I. Saunders, *Failure of simulation training to change residents' management of oesophageal intubation.* British Journal of Anaesthesia, 2003. **91**(3): p. 312-318.

61. Ostergaard, D., *National Medical Simulation Training in Denmark.* Critical Care Medicine, 2004. **32**(2 (suppl)): p. S58-S60.

62. Nyssen, A., R. Larbuisson, M. Janssens, P. Pendeville, and A. Mayne, *A comparison of the training value of two types of anesthesia simulators: computer screen-based and mannequin-based simulators.* Anesthesia and Analgesia, 2002. **94**: p. 1560-5.

63. Helmriech, R., A. Merrit, and J. Wilhelm, *The evolution of crew resource management training in commercial aviation.* International Journal of Aviation Psychology, 1999. **9**: p. 19-32.

64. Howard, S., D. Gaba, J. Fish, G. Yang, and F. Sarnquist, *Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents.* Aviation, Space and Environmental Medicine, 1992. **63**: p. 763-70.

65. Roewer, N., C. Krier, and G. Noldge-Shomburg, *Anesthesia simulation and patient security-from innovation to a standard procedure.* Anasthesiol Intensivmed Notfallmed Schmerzther, 2004. **39**: p. 229-247.

66. Small, S., R. Wuerz, R. Simon, N. Shapiro, A. Conn, and G. Setnick, *Demonstration of high-fidelity simulation team training for Emergency Medicine.* Academic Emergency Medicine, 1999. **6**(4): p. 312-23.

67. Reznek, M., R. Smith-Coggins, S. Howard, K. Kiran, P. Harter, Y. Sowb, D. Gaba, and T. Krummel, *Emergency medicine crisis resource management (EMCRM): pilot study of a simulation-based crisis management course for emergency medicine.* Academic Emergency Medicine, 2003. **10**(4): p. 386-389.

68. McLaughlin, S., D. Doezema, and D. Sklar, *Human simulation in emergency medicine training: a model curriculum.* Academic Emergency Medicine, 2002. **9**(11): p. 1310-1318.

69. Bond, W., L. Deitrick, D. Arnold, M. Kostenbader, G. Barr, S. Kimmel, and C. Worrilow, *Using simulation to instruct emergency medicine residents in cognitive forcing strategies.* Academic Medicine, 2004. **79**(5): p. 438-446.

70. Macedonia, C., R. Gherman, and A. Satin, *Simulation Laboratories for training in obstetrics and gynecology.* Obstetrics and Gynecology, 2003. **102**(2): p. 388-392.

71. Halamek, L., D. Kaegi, D. Gaba, and e. al, *Time for a new paradigm in pediatric medical education: teaching neonatal resuscitation in a simulated delivery room environment.* Pediatrics, 2000. **106**(4): p. E45.

72. Watterson, L., B. Flanagan, B. Donavan, and B. Robinson, *Anaesthetic Simulators: Training for the broader health-care profession.* Australian and New Zealand Journal of Surgery, 2000. **70**: p. 735-7.

73. Chopra, V., J. Gesink, J. De Jong, J. Bovill, J. Spierdijk , and R. Brand, *Does training on an anaesthesia simulator lead to improvement in performance?* British Journal of Anaesthesia, 1994. **73**: p. 293-297.

74. Garden, A., B. Robinson, J. Weller, L. Wilson, and D. Crone, *Education to address medical error - a role of high fidelity patient simulation.* New Zealand Medical Journal, 2002. **115**(1150): p. 133-144.

75. Holzman, R., J. Cooper, D. Gaba, J. Philip, S. Small, and D. Feinstein, *Anesthesia crisis resource management: real-life simulation training in operating room crises.* Journal of Clinical Anesthesia, 1995. **7**: p. 675-687.

76. Lampotang, S., M.L. Good, R. Westhorpe, J. Hardcastle, and R. Carovano, *Logistics of conducting a large scale individual sessions with a full-scale patient simulator at a scientific meeting.* Journal of Clinical Monitoring, 1997. **13**: p. 399-407.

77. Wong, S., K. Ng, and P. Chen, *The application of clinical simulation in crisis management training.* Hong Kong Medical Journal, 2002. **8**: p. 131-5.

78. Sica, G., D. Barron, R. Blum, T. Frenna, and D. Raemer, *Computerized realistic simulation: a teaching module for crisis management in radiology.* American Journal of Radiology, 1999. **177**: p. 301-304.

79. Treloar, D., *On-site and distance education of emergency medicine personnel with a human patient simulator.* Military Medicine, 2001. **166**(11): p. 1003-6.

80. McLennan, B., *Early experience with simulated trauma resuscitation.* Canadian Journal of Surgery, 1999. **1999**(3): p. 205-10.

81. Wilkins, D., *Simulator training in anaesthesia.* Current Opinion In Anesthesiology, 1997. **10**: p. 481-484.

82. Berkenstadt, H., A. Ziv, D. Barusk, I. Levine, A. Cohen, and A. Vardi, *The use of advanced simulation in the training of anesthesiologists to treat chemical warfare casualties.* Anesthesia and Analgesia, 2003. **96**(6): p. 1739-42.

83. Kyle, R., D. Via, R. Lowy, J. Madsen, A. Marty, and P. Mongan, *A multidisciplinary approach to teach responses to weapons of mass destruction and terrorism using combined simulation modalities.* Journal of Clinical Anesthesia, 2004. **16**(2): p. 152-8.

84. Murray, W., M. Good, J. Gravenstein, J. van Oostrom, and W. Brasfield, *Learning about new anesthetics using a model driven, full human simulator.* Journal of Clinical Monitoring and Computing, 2002. **17**: p. 293-300.

85. Dalley, P., B. Robinson, J. Weller, and C. Caldwell, *The use of high-fidelity human patient simulation and the introduction of new anaesthesia delivery systems.* Anesthesia and Analgesia, 2004. **99.**(6): p. 1737-1741.

86. Via, D., R. Kyle, J. Trask, C. Shields, and P. Mongan, *Using high-fidelity patient simulation and an advanced distance education network to teach pharmacology to second-year medical students.* Journal of Clinical Anesthesia, 2004. **16**(2): p. 144-51.

87. Cooper, J., D. Barron, R. Blum, K. Davison, D. Feinstein, J. Halasz, D. Raemer, and R. Russell, *Video Teleconferencing with realistic simulation for medical education.* Journal of Clinical Anesthesia, 2000. **12**: p. 256-261.

88. Glavin, R. and D. Greaves, *The problem of coordinating simulator-based instruction with experience in the real workplace.* British Journal of Anaesthesia, 2003. **91**(3): p. 309-311.

89. Eraut, M. and B. du Boulay, *Developing the attributes of medical professional judgement and competence*, University of Sussex: Sussex.

90. Pangaro, L., *A new vocabulary and other innovations from improving descriptive in-training evaluations.* Academic Medicine, 1999. **74**: p. 1203–1207.

91. Byrne, A., A. Sellen, J. Jones, A. Aitkenhead, S. Hussain, F. Gilder, H. Smith, and P. Ribes, *Title*
Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: a multicentre study. Anaesthesia, 2002. **57**(2): p. 176-9.

92. Gordon, M., *A review of the validity and accuracy of self-assessments in Health Professions Training.* Academic Medicine, 1991. **66**(12): p. 762-769.

93. Weller, J. and A. Woodward, *Continuing Medical Education: What for? How? And how much is it worth.* New Zealand Medical Journal, 2004. **117**(1193): p. U876.

94. Murray, D., J. Boulet, A. Ziv, J. Woodhouse, J. Kras, and J. McAllister, *An acute care skills evaluation of graduating medical students: a pilot study using clinical simulation.* Medical Education, 2002. **36**: p. 833-841.

95. Boulet, J., D. Murray, J. Kras, J. Woodhouse, J. McAllister, and A. Ziv, *Reliability and validity of a simulation-based acute care skills assessment for medical students and residents.* Anesthesiology, 2003. **99**(6): p. 1270-1280.

96. Dillon, G., J. Boulet, R. Hawkins, and D. Swanson, *Simulation in the United States Medical Licensing Examination (USMLE).* Quality and Safety in Healthcare, 2004. **13**(Suppl 1): p. i41-i45.

97. Morgan, P. and D. Cleave-Hogg, *A Canadian simulation experience: faculty and student opinions of a performance evaluation study.* British Journal of Anaesthesia, 2000. **85**(5): p. 779-781.

98. Morgan, P. and D. Cleave-Hogg, *Evaluation of medical students' performance using the anesthesia simulator.* Medical Education, 2000. **34**: p. 42-45.

99. Morgan, P., D. Cleave-Hogg, and C. Guest, *Validity and reliability of undergraduate performance assessments in an anesthesia simulator.* Canadian Journal of Anaesthesia, 2001. **48**(3): p. 225-233.

100. Crossley, J., H. Davies, G. Humphris, and B. Jolly, *Generalisability: a key to unlock professional assessment.* Medical Education, 2002. **36**(10): p. 972-978.

101. Newble, D. and D. Swanson, *Psychometric characteristics of the objective structured clinical examination.* Medical Education, 1988. **22**: p. 325-334.

102. Morgan, P., D. Cleave-Hogg, and C. Guest, *A comparison of global ratings and checklist scores for an undergraduate assessment using an anesthesia simulator.* Academic Medicine, 2001. **76**: p. 1053-5.

103. Morgan, P. and D. Cleave-Hogg, *Comparison between medical students' experience, confidence and competence.* Medical Education, 2002. **36**: p. 534-9.

104. Morgan, P., D. Cleave-Hogg, S. DeSousa, and J. Tarshis, *High-fidelity patient simulation: validation of performance checklists.* British Journal of Anaesthesia, 2004. **92**(3): p. 388-92.

105. Rogers, P., H. Jacob, A. Rashwan, and M. Pinksy, *Quantifying learning in medical students during a critical care medicine elective: A comparison of three evaluation instruments.* Critical Care Medicine, 2001. **29**(6): p. 1268-1273.

106. Gordon, J., D. Tancredi, L. Binder, W. Wilkerson, and D. Shaffer, *Assessment of a clinical performance evaluation tool for use in a simulator-based testing environment: a pilot study.* Academic Medicine, 2003. **78**(10): p. S45-S47.

107. Tsai, T.C., P.H. Harasym, C. Nijssen-Jordan, P. Jennett, and G. Powell, *The quality of a simulation examination using a high-fidelity child manikin.* Medical Education, 2003. **37 Suppl 1**: p. 72-8.

108. Holcomb, J.B., R.D. Dumire, J.W. Crommett, C.E. Stamateris, M.A. Fagert, J.A. Cleveland, G.R. Dorlac, W.C. Dorlac, J.P. Bonar, K. Hira, N. Aoki, and K.L. Mattox, *Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training.* Journal of Trauma-Injury Infection & Critical Care, 2002. **52**(6): p. 1078-85; discussion 1085-6.

109. Byrne, A. and J. Greaves, *Assessment instruments used during anaesthetic simulation: review of published studies.* British Journal of Anaesthesia, 2001. **86**(3): p. 445-450.

110. Forrest, F., M. Taylor, K. Postlethwaite, and R. Aspinall, *Use of a high-fidelity simulator to develop testing of the technical performance of novice anaesthetists.* British Journal of Anaesthesia, 2002. **88**(3): p. 338-344.

111. Murray, D., J. Boulet, J. Kras, J. Woodhouse, T. Cox, and J. McAllister, *Acute care skills in anesthesia practice. A simulation-based resident performance assessment.* Anesthesiology, 2004. **101**(5): p. 1084-95.

112. Schwid, H., A. Rooke, J. Carline, R. Steadman, B. Murray, M. Olympio, S. Tarver, K. Steckner, and S. Wetstone, *Evaluation of Anesthesia residents using mannequin-based simulation.* Anesthesiology, 2002. **97**(6): p. 1434-4.

113. Schwid, H., A. Rooke, P. Michalowski, and B. Ross, *Screen-based anesthesia simulation with debriefing improves performance in a mannequin-based anesthesia simulator.* Teaching and Learning in Medicine, 2001. **13**(2): p. 92-96.

114. Devitt, J., M. Kurrek, M. Cohen, and D. Cleave-Hogg, *The validity of performance assessments using simulation.* Anesthesiology, 2001. **95**(1): p. 36-42.

115. Devitt, J., M. Kurrek, M. Cohen, K. Fish, P. Fish, P. Murphy, and J. Szalai, *Testing the raters: inter-rater reliability of standardized anaesthesia simulator performance.* Canadian Journal of Anaesthesia, 1997. **44**(9): p. 924-928.

116. Devitt, J., M. Kurrek, M. Cohen, K. Fish, P. Fish, A. Noel, and J. Szalai, *Testing internal consistency and construct validity during evaluation of performance in a patient simulator.* Anesthesia and Analgesia, 1998. **86**(6): p. 1157-1159.

117. Byrne, A. and J. Jones, *Responses to simulated anaesthetic emergencies by anaesthetists with different durations of clinical experience.* British Journal of Anaesthesia, 1997. **78**: p. 553-556.

118. Lindekaer, A., J. Jacobsen, G. Anderson, M. Laub, and P. Jensen, *Treatment of ventricular fibrillation during anaesthesia in an anaesthesia simulator.* Acta Anaesthesiologica Scandinavica, 1997. **41**: p. 1280-4.

119. Byrne, A. and J. JG, *Inaccurate reporting of simulated critical anaesthetic incidents.* British Journal of Anaesthesia, 1997. **78**: p. 637-41.

120. Gaba, D., S. Howard, B. Flanagan, and K. Fish, *Assessment of clinical performance during simulated crises using both technical and behavioural ratings.* Anesthesiology, 1998. **89**(1): p. 8-18.

121. Weller, J., M. Bloch, S. Young, M. Maze, S. Oyesola, J. Wyner, D. Dob, K. Haire, J. Durbridge, T. Walker, and D. Newble, *Evaluation of the high fidelity patient simulator in the assessment of performance of anaesthetists.* British Journal of Anaesthesia, 2003. **90**(1): p. 43-47.

122. Regehr, G., H. MacRae, R. Reznick, and D. Szalay, *Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination.* Academic Medicine, 1998. **73**(9): p. 993-997.

123. Jacobsen, J., A. Lindekaer, D. Ostergaard, K. Nielsen, M. Laub, P. Jensen, and N. Johannessen, *Management of anaphylactic shock evaluated using a full-scale anaesthesia simulator.* Acta Anaesthesiologica Scandinavica, 2001. **45**: p. 315-9.

124. i Gardi, T., U. Christensen, J. Jacobsen, P. Jensen, and H. Ording, *How do anaesthesiologists treat malignant hyperthermia in a full-scale anaesthesia simulator?* Acta Anaesthesiologica Scandinavica, 2001. **45**: p. 1032-5.

125. Riley, R., D. Wilks, and J. Freeman, *Anaesthetists' attitudes towards an anaesthesia simulator. A comparative survey: USA and Australia.* Anaesthesia and Intensive Care, 1997. **25**(5): p. 514-519.

126. Davis, D., *Impact of formal continuing medical education: do conferences, workshops, rounds and other traditional continuing education activities change physician behaviour or health care outcomes?* JAMA, 1999. **289**(9): p. 867-874.

127. Thomas O'Brien, M., N. Freemantle, A. Oxman, F. Wolf, D. Davis, and J. Herrin, *Continuing education meetings and workshops: effect on professional practice and health care outcomes*. 2001, The Cochrane Database of Systematic Reviews.

128. Nyman, M., M. Murphy, P. Schryver, J. Naessens, and S. Smith, *Improving performance in diabetes care: a multicomponent intervention.* Effective Clinical Practice, 2000. **3**(5): p. 205-212.

129. Goldberg, H., R. Deyo, V. Taylor, A. Cheadle, D. Conrad, J. Loeser, P. Heagerty, and P. Diehr, *Can evidence change the rate of back surgery? A randomized trial of community-based education.* Effective Clinical Practice, 2001. **4**(3): p. 95-104.

130. Tu, K. and D. Davis, *Can we alter physician behaviour by educational methods? Lessons learned from studies of the management and follow-up of hypertension.* Journal of Continuing Education in the Health Professions, 2002. **22**(1): p. 11-22.

131. Combes, X., B. Le Roux, P. Suen, M. Dumerat, C. Motamed, S. Sauvat, P. Duvaldestin, and G. Dhonneur, *Unanticipated difficult airway in anesthetized patients: prospective validation of a management algorithm.* Anesthesiology, 2004. **100**(5): p. 1146-1150.

132. Davis, D. and A. Taylor-Vaisey, *Translating guidelines into practice. A systematic review of the theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines.* Canadian Medical Association Journal, 1997. **157**(12): p. 1736-1737.

133. Davis, D., M. Thomson, A. Oxman, and R. Haynes, *Changing physician performance: a systematic review of the effect of continuing medical education strategies.* JAMA, 1995. **274**(9): p. 700-705.

134. Heinemann, A., E. Roth, K. Rychlik, K. Pe, C. King, and J. Clumpner, *The impact of stroke practice guidelines on knowledge and practice patterns of acute care health professionals.* Journal of Evaluation in Clinical Practice, 2003. **9**(2): p. 203-212.

135. Ketley, D. and K. Woods, *Impact of clinical trials on clinical practice: example of thrombolysis for acute myocardial infarction.* Lancet, 1993. **342**: p. 891-894.

136. Sanci, L., C. Coffey, F. Veit, M. Carr-Gregg, G. Patton, N. Day, and G. Bowes, *Evaluation of the effectiveness of an educational intervention for general practitioners in adolescent health care: randomised controlled trial.* British Medical Journal, 2002. **320**(7229): p. 224–230.

137. Abouleish, A., A. Golden, A. O'Donnell, P. Beach, K. Calhoun, T. Blackwell, P. Galloway, L. Teske, S. Wilson, E. Koleng, and A. Leblanc, *Problem-based learning in a managed care seminar for all new residents in an academic medical centre.* Texas Medicine, 2003. **99**(2): p. 54-57.

138. Allen, M., J. Sargeant, K. Mann, M. Fleming, and J. Premi, *Videoconferencing for practice-based small-group continuing medical education: feasibility, acceptability, effectiveness, and cost.* Journal of Continuing Education in the Health Professions, 2003. **23**(1): p. 38-47.

139. Mulvey, G., L. Keogh, and M. Temple-Smith, *Outcomes of an educational activity with Victorian GPs aimed at improving knowledge and practices in relation to sexually transmissible diseases.* Australian and New Zealand Journal of Public Health, 2000. **24**(1): p. 76-78.

140. Harrison, C. and W. Hogg, *Why do doctors attend traditional CME events if they don't change what they do in their surgeries? Evaluation of doctors' reasons for attending a traditional CME programme.* Medical Education, 2003. **37**: p. 884-888.

141. de Villiers, M., G. Bresnick, and B. Mash, *The value of small group learning: an evaluation of an innovative CPD programme for primary care medical practitioners.* Medical Education, 2003. **37**: p. 815-821.

142. Freemantle, N., E. Harvey, F. Wolf, J. Grimshaw, R. Grilli, and L. Bero, *Printed educational materials: effects on professional practice and health care outcomes*. 2002, The Cochrane Database of Systematic Reviews.

143. Jamtvedt, G., J. Young, D. Kristoffersen, M. Thomas O'Brien, and A. Oxman, *Audit and feedback: effects on professional practice and health care outcomes.* Cochrane Database of Systematic Reviews, 2003. **3**(CD000259).

144. Thomas O'Brien, M., A. Oxman, R. Haynes, D. Davis, N. Freemantle, and E. Harvey, *Local opinion leaders: effects on professional practice and health care outcomes*. 2001, The Cochrane Database of Systematic Reviews.

145. Holm, A., *Quality issues in continuing medical education.* British Medical Journal, 1998. **316**: p. 621-624.

146. Chambers, W., K. Ferguson, and G. Prescott, *Continuing medical education by anaesthetists in Scotland: activities, motivation and barriers.* Anaesthesia, 2000. **55**: p. 1192-1212.

147. Bashook, P. and J. Parboosingh, *Continuing medical education: recertification and maintenance of competence.* British Medical Journal, 1998. **316**: p. 545-548.

148. Kolb, D.A., *Experiential Learning: Experience as the source of learning and development,* ed. I. Prentice-Hall, Englewood Cliffs, N.J. 1984.

149. Weller, J. and M. Harrison, *Continuing education and New Zealand anaesthetists: an analysis of current practice and future needs.* Anaesthesia and Intensive Care, 2004. **32**(1): p. 59-63.

150. Weller, J., L. Wilson, and B. Robinson, *Survey of change in practice following simulation-based training in crisis management.* Anaesthesia, 2003. **58**(5): p. 471-473.

151. Weller, J., A. Dowell, M. Kljakovic, and B. Robinson, *Simulation training for medical emergencies in general practice.* Medical Education, 2005. **in press**.

152. McQuillan, P., S. Pilkington, A. Allan, and B. Tayler, *Confidential enquiry into quality of care before admission to intensive care.* BMJ, 1998(316): p. 1853-8.

153. Goldhill, D., W. SA, and A. Sumner, *Physiological values and procedures in the 24 hours before ICU admission from the ward.* Anaesthesia, 1999. **54**: p. 529-34.

154. McGloin, H., S. Adam, and M. Singer, *Unexpected deaths and referrals to intensive care of patients on general wards. Are some cases potentially avoidable?* Journal of the Royal College of Physicians of London, 1999. **33**(3): p. 255-9.

155. Schein, R.M., N. Hazday, M. Pena, B.H. Ruben, and C.L. Sprung, *Clinical antecedents to in-hospital cardiopulmonary arrest.* Chest, 1990. **98**(6): p. 1388-92.

156. Franklin, C. and J. Mathew, *Developing strategies to prevent inhospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event.* Critical Care Medicine, 1994. **22**(2): p. 244-7.

157. Neale, G., *Risk management in the care of medical emergencies after referral to hospital.* Journal of the Royal College of Physicians of London, 1998. **32**(2): p. 125-9.

158. Tekian, A., *Have newly graduated physicians mastered essential clinical skills?* Medical Education, 2002. **36**: p. 406-407.

159. Smith, G. and N. Poplett, *Knowledge of aspects of acute care in trainee doctors.* Postgraduate medical journal, 2002. **78**: p. 335-8.

160. Harrison, G., K. Hillman, G. Fulde, and T. Jacques, *The need for undergraduate education in Critical Care. (Results of a questionnaire to Year 6 medical undergraduates, University of New South Wales and recommendations on a curriculum in Critical Care).* Anaesthesia and Intensive Care, 1999. **27**(1): p. 53-8.

161. Remes, V., I. Sinisaari, A. Harjula, and I. Helenius, *Emergency procedure skills of graduating medical doctors.* Medical Teacher, 2003. **25**(2): p. 149-154.

162. Meek, T., *New house officers' knowledge of resuscitation, fluid balance and analgesia.* Anaesthesia, 2000. **55**: p. 1127-43.

163. Stoneham, M., G. Saville, and WilsonIH, *Knowledge about pulse oximetry among medical and nursing staff.* Lancet, 1994. **344**: p. 1339-42.

164. Teahon, K. and D. Bateman, *A survey of intravenous drug administration by preregistration house officers.* BMJ, 1993. **307**: p. 605-6.

165. Montgomery, H., S. Hunter, S. Morris , R. Naunton-Morgan, and R. Marshall, *Interpretation of electrocardiograms by doctors.* BMJ, 1994. **309**: p. 1551-2.

166. Goldacre, M., T. Lamber, J. Evans, and G. Turner, *Preregistration house officers' view on whether their experience at medical school prepared them will for their jobs: a national questionnaire survey.* BMJ, 2003. **326**: p. 1011-2.

167. Buchman, T., R. Dellinger, R. Raphaely, and e. al, *Undergraduate education in critical care medicine.* Critical Care Medicine, 1992. **20**: p. 1595-1603.

168. Frankel, H., P. Rogers, R. Gandhi, E. Freid, O. Kirton, and M. Murray, *What is taught, what is tested: findings and competency-based recommendations of the Undergraduate Medical Education Committee of the Society of Critical Care Medicine.* Critical Care Medicine, 2004. **32**(9): p. 1949-1956.

169. Garcia-Barbero, m. and J. Such, *Teaching critical care in Europe: analysis of a survey.* Critical Care Medicine, 1996. **24**(4): p. 696-704.

170. Cook, C. and G. Smith, *Do textbooks of clinical examination contain information regarding the assessment of critically ill patients?* Resuscitation, 2004. **60**: p. 129-36.

171. Rogers, P., A. Grevnik, and R. Willenkin, *Teaching medical students complex cognitive skills in the intensive care unit.* Critical Care Medicine, 1995. **1995**(23): p. 575-81.

172. DeBehnke, D., D. Shepherd, and J. Ma, *A case-based Emergency Medicine curriculum for senior medical students.* Academic Emergency Medicine, 1994. **2**(6): p. 519-522.

173. Culhane, A., M. Kamien, and A. Ward, *The contribution of an undergraduate rural attachment to the learning of basic practical and emergency procedural skills.* Medical Journal of Australia, 1993. **159**: p. 450-2.

174. David, S., S. Selvaranjini, and M. Thomas, *Incorporation of emergency medicine in the undergraduate curriculum.* The National Medical Journal of India, 1997. **10**(2): p. 80-81.

175. Coates, W., *An educator's guide to teaching emergency medicine to medical students.* Academic Emergency Medicine, 2004. **11**(3): p. 300-6.

176. Celenza, A., G. Jelinek, I. Jacobs, C. Kruk, R. Graydon, and L. Murray, *Implementation and evaluation of an undergraduate emergency medicine curriculum.* Emergency Medicine, 2001. **13**: p. 98-103.

177. Burdick, W., N. Jouriles, G. D'Onofrio, L. Kass, J. Mahoney, and K. Restifo, *Emergency Medicine in undergraduate education.* Academic Emergency Medicine, 1998. **5**(11): p. 1105-9.

178. Kelly, A. and M. Ardagh, *Does learning emergency medicine equip medical students for ward emergencies.* Medical Education, 1994. **28**(524-527).

179. Smith, G., V. Osgood, and S. Crane, *ALERT a multiprofessional training course in the care of the acutely ill patient.* Resuscitation, 2002. **52**: p. 281-6.

180. Smith, G.B. and N. Poplett, *Impact of attending a 1-day multi-professional course (ALERT) on the knowledge of acute care in trainee doctors.* Resuscitation, 2004. **61**(2): p. 117-22.

181. Girard, M. and P. Drolet, *Anesthesiology simulators: networking is the key*. Canadian Journal of Anaesthesia, 2002. **49**(7): p. 647-649.

182. Weller, J., *Simulation in undergraduate medical education: bridging the gap between theory and practice.* Medical Education, 2004. **38**: p. 32-38.

183. Weller, J., B. Robinson, P. Larsen, and C. Caldwell, *Simulation-based training to improve acute care skills in medical undergraduates.* New Zealand Medical Journal, 2004. **117**(1204).

184. Lawson, K., *The Trainer's Handbook*. 1998, New York: Jossey-Bass, Pfeiffer.

185. Brookfield, S., *Understanding and Facilitating Adult Learning*. 1986, Milton Keynes: Open University Press.

186. Prideaux, D., *Researching the outcomes of educational interventions: a matter of design.* British Medical Journal, 2002. **324**: p. 126-127.

187. Smits, P., J. Verbeek, and C. de Buisonjé, *Problem based learning in continuing medical education: a review of controlled evaluation studies.* British Medical Journal, 2002. **324**: p. 153-156.

188. Weller, J., B. Robinson, B. Jolly, L. Watterson, M. Joseph, S. Bajenov, A. Houghton, and P. Larsen, *Psychometric characteristics of simulation-based assessment in anaesthesia and accuracy of self-assessed scores.* Anaesthesia, 2005. **60**(3): p. 245-50.

189. Van der Vleuten, C., *The assessment of professional competence: development, research and practical implications.* Advances in Health Sciences Education, 1996. **1**: p. 41-67.

190. Miller, G., *Performance assessment.* Academic Medicine, 1990. **65**(9): p. S63-S67.

191. Norman, G., G. Bordage, L. Curry, D. Dauphinee, B. Jolly, D. Newble, A. Rothman, B. Stalenhoef, P. Stillman, D. Swanson, and X. Tonesk, *A review of recent innovations in assessment*, in *Directions In Clinical Assessment*, R. Wakeford, et al., Editors. 1995.

192. Norcini, J., L. Blank, and F. Duffy, *The mini-CEX: A method for assessing clinical skills.* Annals of Internal Medicine, 2003. **138**: p. 476-481.

193. Wass, V., R. Jones, and C. Van der Vleuten, *Standardized or real patients to test clinical competence? The long case revisited.* Medical Education, 2001. **35**: p. 321-325.

194. Norcini, J., D. Swanson, and G. Webster. *Reliability, validity and efficiency of various item formats in assessment of physician competence*. in *Proceedings of the 22nd Annual Conference on Research in Medical Education*. 1983. Washington DC.

195. Norman, G., L. Muzzin, R. Williams, and D. Swanson, *Simulation in health science education.* Journal of Instructional Development, 1985. **8**: p. 11-17.

196. Stillman, P., D. Swanson, S. Smee, A. Stillman, T. Ebert, and V. Emmel, *Assessing clinical skills in residents with standardized patients.* Annals of Internal Medicine, 1986. **105**: p. 762-771.

197. Gray, J., *Global ratings in residency education.* Academic Medicine, 1996. **71**(Supplement 1): p. S55-S63.

198. Boud, D., *Enhancing Learning Through Self Assessment*. 1 ed. 1995, London, Philadelphia: Kogan Page. 247.

199. DeAnda, A. and D. Gaba, *Role of experience in the response of simulated critical incidents.* Anesthesia and Analgesia, 1991. **72**: p. 308-315.

200. MacDonald, J., R. Williams, and D. Rogers, *Self-assessment in simulation-based surgical skills training.* American Journal of Surgery, 2003. **185**: p. 319-322.

201. Martin, D., G. Regehr, B. Hodges, and N. McNaughton, *Using videotaped benchmarks to improve the self-assessment ability of Family Practice residents.* Academic Medicine, 1998. **73**(11): p. 1201-1206.

202. Gordon, J., *Self-assessment programs and their implications for Health Professions Training.* Academic Medicine, 1992. **67**(10): p. 672-679.

203. Smith, S. and E. Vela, *Environmental context-dependent memory: a review and meta-analysis.* Psychon Bull Rev. . 2001. **8**(2): p. 203-20.

204. Godden, D. and A. Baddely, *Context-dependent memory in two natural environments on land and underwater.* British Journal of Psychology, 1975. **66**(3): p. 325-331.

205. Ramsden, P., *Improving learning. New Perspectives*. 1988: Kogan Page.